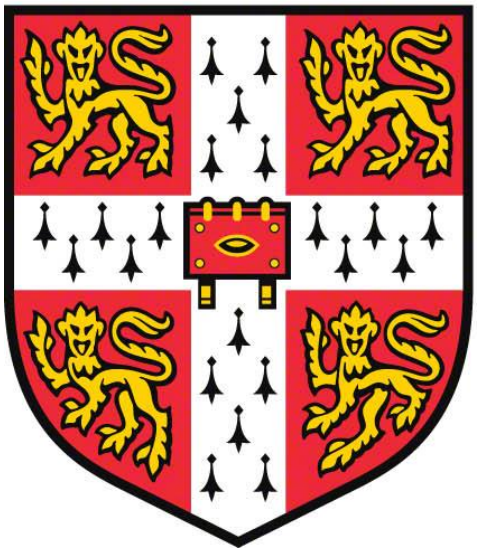
The effect of replication impediments on differentiation

Cara Bernadette Eldridge
University of Cambridge
St Catharine's College
March 2019



This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Abstract

Cara Bernadette Eldridge

The effect of replication impediments on differentiation

In this thesis I set out to answer the question 'is differentiation robust to replication impediments?'. Prior work in the group has focussed on the epigenetic impact of replication impediments in the terminally differentiated DT40 chicken cell line. I wanted to find out whether these impediments could impact the fundamental biological process of embryonic development. In this work the BOBSC human induced pluripotent stem cell line was differentiated to definitive endoderm while replication was perturbed using DNA damaging agents, low dose hydroxyurea and G-quadruplex secondary structure stabilisation.

DNA damage was induced both in the undifferentiated state and during differentiation. When damage was induced during differentiation, greater levels of cell death were seen and there was a larger increase in cells in G2/M phase of the cell cycle compared with treatment in the undifferentiated state. The efficiency of differentiation was observed to negatively correlate with the dose of DNA damaging agent.

During unperturbed differentiation, the levels of DNA damage response proteins, including p53, were found to decrease. However, when DNA damage was induced during differentiation the level of p53 increased. In order to understand whether the upregulation of p53 was preventing differentiation, *TP53*$^{-/-}$ cells were differentiated in the presence of DNA damaging agents. These cells were found to differentiate as efficiently as wildtype untreated cells. This indicates that p53 prevents differentiation in the presence of DNA damage, termed here a 'differentiation checkpoint'.

The level of γH2A.X was found to be markedly increased 50 hours into unperturbed differentiation. This correlates with the point at which the cells transition from being epithelial to mesenchymal. The role of this rise is not known but may correspond to massive transcriptional changes that occur during this transition.

G-quadruplex-binding ligands were also shown to alter the course of differentiation, either by a p53-dependent mechanism or by a separate, possibly G-quadruplex specific, mechanism. This also occurred in *REV1*$^{-/-}$ cells, a specialised polymerase known to play a role in G-quadruplex-processing.

This work has led to the conclusion that replication impediments are able to alter the course of definitive endoderm differentiation in human induced pluripotent stem cells.

Abbreviations

Abbreviation	Meaning
2i	Two inhibitor conditions: GSK3i and MEKi
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
6-4PP	6-4 photoproducts
AEJ	Alternative end joining
ATM$^{-/-}$	*ATM* deficient cell line
BER	Base excision repair
bp	Base pairs
cDNA	Complementary DNA
ChIPseq	Chromatin immunoprecipitation sequencing
CPD	Cyclobutane-pyridmidine dimers
DAPI	4′,6-diamidino-2-phenylindole
DDR	DNA damage response
DE	Definitive endoderm
DMSO	Dimethylsulfoxide
DSB	DNA double strand break
dsDNA	Double stranded DNA
EB	Elution buffer (Qiagen kit)
EdU	5-ethynyl-2'-deoxyuridine
eGFP	Enhanced green fluorescent protein
EMT	Epithelial mesenchymal transition
EpiLC	Epiblast-like stem cell
EpiSC	Epiblast stem cell
G4	G-quadruplex
gDNA	Genomic DNA
GO	Gene ontology
GSK3i	Glycogen synthase kinase 3 inhibitor
HDAC	Histone deacetylase
hESC	Human embryonic stem cell
hiPSC	Human induced pluripotent stem cell
HR	Homologous recombination

HU	Hydroxyurea
ICM	Inner cell mass
KO	Knockout
LIF	Leukaemia inhibitory factor
MEF	Mouse embryonic fibroblast
MEKi	Mitogen-activated protein kinase inhibitor
mESC	Mouse embryonic stem cell
MET	Mesenchymal epithelial transition
MMEJ	Microhomology-mediated end joining
MMS	Methyl methanesulfonate
mtDNA	Mitochondrial DNA
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NMM	N-methyl mesoporphyrin IX
TP53$^{-/-}$	*TP53* deficient cell line
PCA	Principle component analysis
PDS	Pyridostatin
PhenDC3	3,3′-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate (1:2)
PI	Propidium iodide
PMSF	Phenylmethane sulfonyl fluoride
Pol II	RNA polymerase II
PRIMPOL$^{-/-}$	*PRIMPOL* deficient cell line
PS	Primitive streak
PSD	Pluripotent state dissolution
PTM	Post-translational modification
qPCR	Quantitative PCR
REV1$^{-/-}$	*REV1* deficient cell line
RNAseq	RNA sequencing
RNR	Ribonucleotide reductase
ROS	Reactive oxygen species
RT	Room temperature
SCID/beige	Severe combined immunodeficient beige
SPR	Surface plasmon resonance

ssDNA	Single stranded DNA
TF	Transcription factor
TSS	Transcription start site
UD	Undifferentiated
UFB	Ultrafine anaphase bridge
UT	Untreated
UTR	Untranslated region
UV	Ultraviolet light
WRN$^{-/-}$	*WRN* deficient cell line
WT	Wildtype

Acknowledgements

I would like to thank Julian Sale and all members of the Sale lab for help and advice throughout the project. Julian has been incredibly supportive throughout this project, always happy to listen to new ideas and give interesting feedback. He has always kept me upbeat and been excited by this project. I would like to thank all members of the lab, past and present, especially Benedicte Recolin for amazingly patient daily help when desperately needed. I would like to thank Hayat Arzouk for help with flow cytometry and RNA sequencing when I was starting out, Marina Romanello and Davide Schiavone for qPCR set up and analysis help as well as ChIP. I am grateful for Leticia Koch-Lerner's input into my thesis with her vast depth of DDR knowledge as well as help with the fragment analysis. Leah Parry was a great summer student who also joined me in the project for an eight-week placement. I would like to thank Gareth Williams for useful discussion and scientific help throughout this project. I am grateful for Sasa Svikovic's help with phenol-chloroform extraction and library preparation. Daniela Peris helped with my western blots and was always around to keep my spirits up. I would like to thank Alastair Crisp for help with sequencing data, PCAs and generation of the PEAT software which was used throughout this thesis. Pierre Murat has also helped with sequencing data analysis. I am thankful for all other members of the lab for input: Guillaume Guilbaud, Chris Gilmartin, Chris Mellor, Jovan Traparic and Lucy Walker. I would also like to thank the labs around for the borrowing of equipment and antibodies, and contributing to the LMB environment.

I am very grateful to my second supervisor, Gerry Crossan for help during the initial year of my project with the mouse ES cells, as well as Richard Pannell for help with cell culture. I am very appreciative of Ludovic Vallier, and a posdoc in his lab Rodrigo Grandy, for guidance throughout the project and in setting up the hiPSCs. I am also very grateful for the help of the LMB flow cytometry facility throughout this project. I thank Madan Babu for interesting discussions regarding my sequencing data.

I would like to thank the MRC and LMB for my funding as well as St Catharine's for additional conference funding.

Finally, I would like to thank Fin Allen for constant support and encouragement every day of my PhD. He not only made sure that I worked as normal as possible hours but would also come into the lab with me on weekends when I was struggling to go in. I am also grateful for his help

with data analysis, including help with derivation of equations, general help with coding and specifically python. He also helped with generation of scripts to convert data to tab delineated binary data, conversion of Ensembl to Entrez ID, Venn diagram scripts and intersections of lists. I would also like to thank all of my family and friends for helping throughout these last three years.

Contents

Chapter 1. Introduction

The aim of this project was to understand whether replication impediments could alter differentiation. This was particularly novel as all work performed in the group previously had utilised a terminally differentiated chicken cell line. Throughout this thesis embryonic stem cells and induced pluripotent stem cells were differentiated in the presence of replication impediments.

1.1 Embryonic stem cells and induced pluripotent stem cells

Both human and mouse embryonic stem cells have been studied in great detail due to their value in medicine, their ability to generate genetic knockout mouse cell lines, and for insight into development. Induced pluripotent stem cells have been and will be incredibly important in regenerative medicine. For this reason, they were considered an ideal system to study in this thesis.

1.1.1 Mouse embryonic stem cells

The culture and differentiation of mouse embryonic stem cells (mESCs) has greatly impacted molecular biology and this made the system a valuable tool for use in this project. During development the zygote becomes the morula and then the blastocyst, prior to implantation and lineage specification. mESCs are derived from the inner cell mass (ICM) of an E3.75-4.75 mouse blastocyst (Boroviak et al., 2014) in the preimplantation epiblast (Evans and Kaufman, 1981; Martin, 1981). The first mESCs cultured *in vitro* were shown to be capable of differentiating *in vitro* as well as forming tumours in an *in vivo* context (*ibid*). These cells are able to replicate indefinitely without losing cell fate, as well as differentiate into the three somatic lineages and germ cells (*ibid*). They are therefore able to form chimaeras contributing to all lineages after *in vitro* cell culture once returned to the epiblast (*ibid*). These cells have been instrumental in the production of genetically engineered mice, more recently using the CRISPR/Cas9 technology (Jinek et al., 2012), and have allowed us to model many diseases including cancer (reviewed by Zhang et al., 2011a).

1.1.1.1 Generation of mESCs in culture

The first embryonic stem cells with a normal karyotype were generated in 1981 using the mouse background 129 SvE (Evans and Kaufman, 1981). However, the generation from other mice was not as straightforward, possibly due to differences in the genetic background (Gardner and Brook, 1997). Furthermore, aside from in human cells, ES cells were not generated from other

animals until 2008 when rat stem cells were generated in culture upon an optimisation of culture conditions, discussed later (Buehr et al., 2008; Li et al., 2008). This also allowed isolation in two other mouse strains C57/BL6 (Kiyonari et al., 2010) and NOD (Nichols et al., 2009). Interestingly, the majority of mES cell lines created are of male XY genotype; this is postulated to be due to the process of X chromosome inactivation and those cells with XX genotype can result in deletions of an X chromosome (Rastan and Robertson, 1985).

As these cells were not common to all organisms, it was a concern that they were an *in vitro* artefact that had no biological relevance. There are two reasons why these cells may be easier to generate in rodents. Firstly, rodents form an egg cylinder (Copp, 1979): after the blastocyst is formed, the epiblast cells become organised into cup-shaped epithelium by apoptosis of the internal epiblast cells and this may facilitate isolation of these cells (Coucouvanis and Martin, 1999). The other concern that is specific to these cells is that rodents are able to enter a state of facultative diapause during differentiation: this state is very specific to their development and is not thought to be shared with primates (reviewed by Paria et al., 2002). This allows embryonic development to enter a state of dormancy prior to implantation to cope with external conditions, for example when a mother is still weaning pups from a previous litter. It can also be induced in an experimental context by lowering levels of oestrogen in the mouse, again enhancing the likelihood of capturing this cell state (Paria et al., 1998).

1.1.1.2 Signalling pathways maintaining the pluripotent state

The initial mESCs were cultured with mouse embryonic fibroblasts (MEFs), later discovered to produce leukaemia inhibitory factor (LIF) (Smith et al., 1988; Williams et al., 1988). LIF acts through STAT3, which is a transcription factor that inhibits differentiation pathways and promotes cell viability (Boeuf et al., 1997; Smith, 2001). LIF binds to the glycoprotein 130 (gp130) cell surface receptor, activating the JAK/STAT3 pathway. This results in phosphorylation of STAT3, causing translocation from the cytoplasm to the nucleus where it binds target DNA. In conventional mESC cultures grown in serum, BMP4 (bone morphogenic protein) binds to BMPR1 and BMPR2 membrane proteins, causing activation of the ERK/MAPK and SMAD signalling pathways and results in SMAD proteins, translocating into the nucleus and activating *Id* (inducer of differentiation) genes (Figure 1) (Ying et al., 2003). The combination of LIF and BMP cause upregulation of pluripotency genes and this combination of factors allows serum-free culture of mESCs (Ying et al., 2003). However, cultures are frequently maintained in serum and LIF and these tend to be heterogeneous in their gene expression of a number of pluripotency genes including *Zfp42, Dppa3, Nanog, Esrrb* and

Klf4: this was not therefore an optimum culturing procedure (van den Berg et al., 2008; Chambers et al., 2007; Hayashi et al., 2008; Niwa et al., 2009; Toyooka et al., 2008).

FGF4 stimulation of the ERK/MAPK signalling pathway primes mESCs for lineage specification, and therefore differentiation (Kunath et al., 2007; Stavridis et al., 2007): the combination of LIF and BMP restrict but do not block this pathway. Inhibition of glycogen synthase kinase 3 (GSK3) stimulates WNT signalling, causing β-CATENIN to relocate to the nucleus, removing inhibition of specific transcription factors and activation of the pluripotency network (Figure 1) (Wray and Hartmann, 2012). Inhibition of the MAPK pathway along with inhibition of GSK3 can be done using two small molecule inhibitors, known as 2i: MEK by PD0325901 and GSK3 by CHIR99021 (Silva and Smith, 2008). This combination is sufficient to keep cells in the naïve pluripotent state with full pluripotency, and this revolutionised culture of these cells (Silva and Smith, 2008). While mESCs can be maintained in serum conditions, these populations are often heterogeneous in morphology and gene expression (Hayashi et al., 2008). However, using the 2i system, growing cells on gelatin coated plates with LIF, the dependence on serum and feeder cells are removed as well as the heterogeneity (Ying et al., 2008).

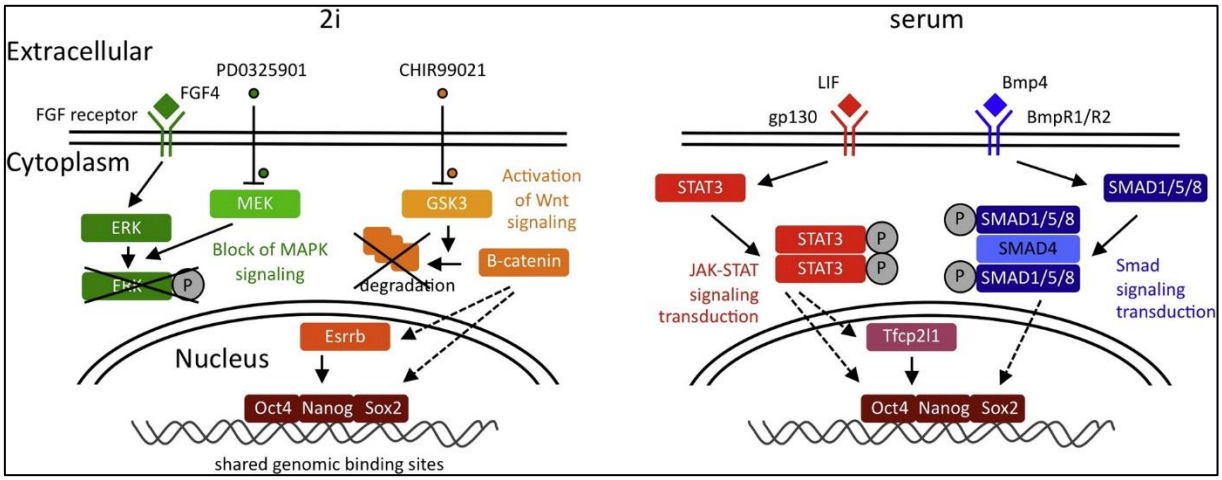


Figure 1. mESCs grown on 2i are more homogenous in nature (left diagram) in comparison to those grown in serum (right). PD0325901 works by inhibiting MEK and stopping MAPK signalling. CHIR00921 inhibits GSK3 and causes β-catenin to build up where it activates its targets in the nucleus. Diagram reproduced from Marks and Stunnenberg, 2014.

1.1.1.3 Gene expression profiles of mESCs

Work from the group of Austin Smith has shown that over a quarter of expressed genes show at least two-fold differences in expression between 2i and serum culture conditions (Marks et al., 2012). However, the expression profile of these cells can be altered simply by switching the culture conditions, so these changes are reversible. Importantly, lineage specification genes

remain downregulated in cells cultured in 2i conditions. Most genes considered to be critical for pluripotency are expressed similarly between 2i and serum cultured mES cells, including *Pou5f1, Sox2, Nanog, Zfp42, Klf2* and *Klf4* (Marks et al., 2012). The OCT4, SOX2, NANOG trio share binding sites in the genome and are key to controlling expression of pluripotency genes, including themselves. Highly expressed pluripotency transcripts in naïve mES cells also include *Dppa3, Tbx3, Gbx2* and *Nodal* (Guo et al., 2009).

Heterogeneity in stem cell populations can be monitored in much greater detail than the initial experiments using single-cell RNA sequencing. This allows separation on the individual cell level of transcriptomes, and has been used in human and mouse ESCs as well as induced pluripotent stem cells (Kolodziejczyk et al., 2015; Nguyen et al., 2018), for example to show differences in naïve and primed human ES cells (Messmer et al., 2019).

1.1.2 Epigenetic state of mESCs

The chromatin in the nucleus holds epigenetic information as well as the genetic code. The epigenetic information is encoded in histone modifications, DNA methylation and non-coding RNA. Epigenetic details are stored in the histones around which the DNA is wrapped, the histone modifications and particular histone variants can broadly lead to euchromatic and heterochromatic regions in the genome. Methylation of the DNA itself, mostly at CpG sites, also causes a change in chromatin structure and alters gene expression. Both of these codes have been shown to play a role in maintenance of the ground state, and specification to differentiate (Sim et al., 2017; Zhang et al., 2012; Zheng et al., 2016).

1.1.2.1 Histone modifications in mESCs

Histones are composed of an $H3_2H4_2$ tetramer and two H2AH2B dimers in which the DNA is wrapped around to form a nucleosome (Arents et al., 1991). These histones can be modified, often on the N-terminal histone tail, to alter the packing of DNA. Modifications include methylation, phosphorylation, sumoylation, ubiquitylation and acetylation. Historically, histone post-translational modifications (PTMs) were considered to be active and associated with euchromatin, for example acetylation, or repressive and linked to heterochromatin formation, such as methylation. This is now known not to be the case and it is clear that there is also cross-talk between marks and dependency upon the location in the genome as to the outcome of the mark (reviewed by Bannister and Kouzarides, 2011).

Histone modifications were initially thought to be key to transcriptional activation in mES cells. Bivalent promoters carrying both the transcriptional activation mark histone 3 lysine 4 trimethylation (H3K4me3) and the repressive mark histone 3 lysine 27 trimethylation (H3K27me3) have attracted much attention due to the idea that these promoters are poised for up or downregulation (Azuara et al., 2006; Bernstein et al., 2006). These bivalent regions have been shown to be high in mES cells cultured in serum but much lower in those cultured in 2i (Marks et al., 2012). The H3K27me3 mark is laid down by polycomb repressive complex 2 (PRC2) and it is this mark that is reduced at promoters of cells grown in 2i compared to those in serum, although the global level is similar (Marks et al., 2012). The H3K4me3 peaks are similar in their frequency and intensity across the genome in serum and 2i cultured cells. Since cells cultured in both ways have a high potential to differentiate, this suggests that bivalent promoters do not exist to poise the cell for differentiation as initially proposed.

RNA polymerase II (Pol II) proximal pausing in collaboration with histone marks has also been identified as a mechanism to control gene expression. Levels of Pol II at the transcription start site (TSS) were higher in 2i cells compared to serum cells (Marks et al., 2012). This mechanism may act to control gene expression quickly, for example in lineage specification, instead of using bivalent chromatin.

1.1.2.2 DNA methylation in mESCs

DNA methylation is a mark that occurs by covalent transfer of a methyl group to the C5 of a cytosine at CpG sites: this transfer is catalysed by DNA methyltransferases (DNMTs). This modification can change the expression of a region of DNA: methylation over the promoter region is generally thought to repress expression of that gene. This modification is known to be required for normal mammalian development, including the process of X-chromosome inactivation and repression of transposable elements (Jones and Takai, 2001).

DNMT3A and DNMT3B primarily set up these modifications and DNMT1 is responsible for maintaining them on the DNA (Okano et al., 1999; Robert et al., 2003). The removal of these marks is catalysed by the TET enzymes, specifically TET1 and TET2 in mES cells. They catalyse the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and finally to unmethylated cytosine (Ficz et al., 2011). This methylation machinery is required for lineage specification and mice lacking these DNMTs do not develop normally, highlighting their importance (Siegfried and Cedar, 1997; Smith and Meissner, 2013). p53 is thought to play

a role in the control of the DNA methylation (Tovy et al., 2017) and this will be discussed further in the DNA damage response section 1.3.

There is also known to be a relationship between the DNA methylation and the histone modifications. This has been shown to occur by proteins known to bind to methylated DNA, such as KAISO, MECP2 and MBD1, which then recruit complexes containing histone deacetylases (HDACs) and histone methyltransferases (reviewed by Bird, 2002). Thereby the histone modifications are altered due to the DNA methylation.

The E3.5 ICM of developing mice has a very low level of DNA methylation, around 20% of CpG islands (Smith et al., 2012). There is a significant difference in the level of DNA methylation between mESCs grown in serum and 2i: serum mESCs are hypermethylated to around 80% whereas those grown on 2i more accurately represent the E3.5 stage in the embryo, having very little DNA methylation (Stadler et al., 2011). Within ten days of switching the culture medium, cells grown on 2i or serum can interconvert between the states in morphology, transcriptome and DNA methylation (Habibi et al., 2013; Marks et al., 2012).

In 2i cells, the DNA methylation occurs in specific euchromatic regions of the genome, and these locations are marked with the histone modification H3K9me3 (Habibi et al., 2013). This mark is thought to be written and sustained by DNMT1, a *de novo* DNA methyltransferase, which is recruited to these regions by UHRF1, and in turn binds to H3K9me3 which is enriched at imprinted control regions and intracisternal A-particle elements (Rothbart et al., 2012).

1.1.3 Cell cycle in mES cells

The cell cycle of mES cells is very fast, measured to be between 10 and 14 hours, with around 65 percent of cells being in S phase at any one time (Pauklin et al., 2011). Only around 15 percent of cells are in G1 phase and it has been suggested that the G1 phase of cells cultured in 2i is shorter than that for cells cultured in serum (Malashicheva et al., 2012). Interestingly, these mES cells do not have a G1/S phase checkpoint, mentioned in more detail with hESCs. This is thought to be due to the lack of Cyclin D, hyperphosphorylation of the retinoblastoma protein and unresponsiveness to CDK4 (Savatier et al., 1996). It is likely that this governs naïve pluripotency. It has been suggested that this short G1 phase causes constitutive replication stress and these cells have upregulated mechanisms to deal with this (Ahuja et al., 2016). Replication stress is defined as the slowing or stalling of replication fork progression, this is discussed in more detail in the DNA damage response section 1.3. However, more recent work has shown

that mES cells cultured in the LIF 2i state have differing cell cycle properties compared to cells grown in serum (ter Huurne et al., 2017). These cells are thought to have an active G1 checkpoint due to hypophosphorylation of Rb and an increased G1 cell cycle phase. Therefore, this cell cycle distribution may not be required for naïve pluripotency.

1.1.4 Epiblast stem cells

After implantation, mESCs cannot be derived from the blastocyst. However, a different cell type can be isolated from E5.5-7.5 mice: post-implantation epiblast-derived cells (EpiSCs) (Tesar et al., 2007). This state no longer represents naïve pluripotency but a state of primed pluripotency in which the cells differ from mESCs in terms of their morphology, growth factor requirements, gene expression profiles, level of DNA methylation and X chromosome activation (Nichols and Smith, 2009). Female mES cells have gone through the process of X chromosome inactivation by the EpiSC stage.

1.1.4.1 Derivation of EpiSCs

EpiSCs have been derived from the post-implantation epiblast of mice by culturing with Activin and FGF but without LIF (Tesar et al., 2007). While these cells can be maintained in the primed state indefinitely, they are not able to contribute to blastocyst chimeras but are able to generate teratomas with differentiated cell types (Guo et al., 2009; Tesar et al., 2007). These cells are able to efficiently differentiate *in vitro* showing that they have a wide differentiation potential (Tesar et al., 2007).

1.1.4.2 Signalling pathways controlling EpiSCs

The signalling pathways maintaining pluripotency in EpiSCs have been shown to be functionally distinct from that of ESCs. mESCs use LIF to signal through the gp130/LIF receptor, described above, to maintain pluripotency (Boeuf et al., 1997). However, in EpiSCs, blocking STAT3 phosphorylation at tyrosine-705 using a JAK inhibitor supported the undifferentiated state, showing that this mechanism is not the same in EpiSCs (Tesar et al., 2007). Inhibition of type I Activin receptor-like kinases 4, 5 and 7 showed that the Activin/Nodal pathway is instead required for maintaining the pluripotent state in EpiSCs. Inhibition of this pathway promoted differentiation of EpiSCs towards neuroectodermal fates (Tesar et al., 2007). This dependence on Activin/Nodal and SMAD2/3 signalling is similar to the pathways controlling pluripotency in human embryonic stem cells, discussed in Section 1.1.5.

1.1.4.3 Gene expression profiles of EpiSCs

Pou5f1, Sox2 and *Nanog* have been shown to be expressed at equivalent levels in naïve and primed mES cells (Tesar et al., 2007). However, a number of transcripts have been shown to have decreased expression on the transition to EpiSCs including *Pecam1, Tbx3, Dppa3, Zfp42* and *Gbx2*, allowing monitoring of this change of state. EpiSCs also exhibit increased expression of transcripts associated with the epiblast and lineage specification genes including *Otx2, Eomes Foxa2, T, Gata6, Sox17, Fgf5* and *Cer1* (Guo et al., 2009; Tesar et al., 2007). Therefore, this change can be tracked in cell culture conditions.

1.1.4.4 Chromatin state of EpiSCs

After implantation of the embryo, at the transition from naïve to primed mESCs in culture, a major wave of DNA methylation occurs such that there is an increase to around 70% methylation at E6.5 (Smith et al., 2012). This is thought to play a role in lineage specification as DNA methylation has been shown to occur at promoters of pluripotency factors (Thiagarajan et al., 2014). Bisulfite sequencing data has shown that the methylation profiles of 2i cells very closely represent E3.5 hypomethylation whereas serum cells mimic the hypermethylation of the E6.5 state (Smith et al., 2012). Less is known about the histone modifications in the EpiSC state but H3K9me3 is thought to act in concert with DNA methylation in silencing in these cells (Tosolini et al., 2018).

1.1.4.5 Differentiation of mESCs to mEpiSCs

mESCs cultured in the 2i state can be differentiated *in vitro* to EpiSCs, by removal of LIF and 2i and the addition of FGF and Activin, these cells are often referred to as epiblast-like cells (EpiLC) (Guo et al., 2009). These differentiated cells imitate the E6.5 mouse epiblast from the post-implantation embryo. In these cells one X chromosome is inactivated in XX karyotypes, they have lost the ability to produce germ cells and there is a change in the expression of transcription factors and cell surface markers, discussed above. This EpiSC mouse state is more representative of the human ESC state suggesting that human ESCs are in a primed state of pluripotency. While the transition from the naïve to primed state is relatively straightforward, it is irreversible and the reversal requires reprogramming (Guo et al., 2009). This transition has been suggested to be more of a state change and not differentiation as such (Vallier unpublished).

1.1.5 Human embryonic stem cells and induced pluripotent stem cells

The first human ESCs (hESCs) derived in 1998 portrayed the desired characteristics of a primate ES cell: cultivation from the preimplantation or peri-implantation embryo, prolonged proliferation in the undifferentiated state and the ability to form of all three embryonic germ layers even after lengthy culture (Thomson et al., 1998). These hESCs were derived from human embryos generated for IVF, after approval. The cells were cultured to the blastocyst stage and the ICMs were removed. They were initially cultured on a layer of mouse embryonic fibroblasts (MEFs) but were able to be transferred to coated plates without a feeder layer.

Three of the cultured cell lines had XY karyotype and the other two had XX karyotype, all were able to be frozen and thawed. The H9 XX cell line was kept in culture for 32 passages and maintained a normal genotype over this time, is still commonly used in cell culture and was used in this project (Thomson et al., 1998). The cell lines produced had high levels of telomerase expression and expressed the cell surface markers stage-specific embryonic antigen SSEA-3, SSEA-4, TRA-1-60, TRA-1-81 and alkaline phosphatase that characterise undifferentiated nonhuman primate ESCs (Thomson et al., 1995).

The Niakan group received approval to continue human embryonic development for longer than ever before, a maximum of 14 days after development, and to use CRISPR/Cas9 in human embryos. This came with the aim of understanding the processes required during human development and the differences between this and mouse development. However, the definitive confirmation of an ESC requires the knowledge that these cells are able to contribute to the germ line, using chimeras. Clearly this function is not possible to test in human cells as it was in the mouse, due to ethical reasons, so this rigorous definition cannot be applied. While these cells have been termed embryonic stem cells, they are different from mESCs in their cell culture requirements, morphology and expression profile. hESCs and mouse EpiSCs share a dependence on Activin/Nodal signalling, flattened morphology and a limited capacity for colonising preimplantation embryos (James et al., 2006). This suggests that these isolated cells represent a later stage of differentiation, analogous with EpiSCs, and this corresponds with the last pluripotent stage before gastrulation (Brons et al., 2007; Greber et al., 2010; Tesar et al., 2007). This, again, highlights the fact that true embryonic stems cells may be a feature specific to rodents.

Instead of monitoring the ability to form chimeras in the cells, these cells were tested for their propensity to produce teratomas, as was previously the method in mESCs. All of the derived

hES cell lines produced teratomas after injection into severe combined immunodeficient beige (SCID/beige) mice. The cells also differentiated *in vitro* when grown without a layer of mouse embryonic fibroblasts with and without LIF (Thomson et al., 1998).

The degree of similarity of these stem cells to the cells they originate from in the embryo is much lower than that of the mouse, raising questions about the culture of these cells. Significant changes in the global DNA methylation was seen between the *in vitro* and *in vivo* cells (Guo et al., 2014; Smith et al., 2014).

The similarity of different hES cell lines to one another has been heavily debated, as well as the differences inflicted using different culture conditions. Different human ES cells have varying propensities of over 100 fold to differentiate towards different lineages (Osafune et al., 2008). The HUES 8 cell line is more inclined towards pancreatic differentiation whereas HUES 3 is best for cardiomyocyte differentiation (Osafune et al., 2008). These differences were suggested to be due to epigenetic changes in these cell lines, as the genetic changes could not account for this (*Ibid*). These differences are clearly very important in differentiation protocols but will also impact the significance of research on some of these specific cell types.

1.1.5.1 Gene expression and signalling in hESCs

Regulatory pathways are poorly understood in hES cells compared to mES cells, but these cells are thought to resemble the EpiSC state in mice (Brons et al., 2007; Tesar et al., 2007). As in mESCs, the trio of pluripotency factors OCT4, SOX2 and NANOG control pluripotency (Wang et al., 2012). However, there are a range of homologs of key transcriptional regulators in mESCs such as TBX3 and ESRRB which are not expressed in hESCs, whereas other factors shown to be dispensable in the mouse system, PRDM14, FOXO1 and LSD1 have been identified to regulate hES cell pluripotency (Adamo et al., 2011; Chia et al., 2010; Zhang et al., 2011b). However, large scale transcriptome analysis has suggested that 75% of the genes expressed in mESCs were also expressed in hESCs (Hirst et al., 2006). Clearly the mouse and human ES cells are very different in the context of the regulation by their signalling pathways and care should be taken when comparing the two systems.

The epigenetics of hESCs have been studied in much less detail than mESCs. However, the role that epigenetic changes play during the differentiation have been identified in greater detail.

1.1.5.2 Reprogramming

The ability to reprogram a cell of any fate into an embryonic stem cell has clearly been an attractive topic due to the potential to solve numerous medical problems, as well as for biological insight. The notion of reprogramming was shown to be correct in John Gurdon's landmark experiments in *Xenopus laevis* (Gurdon, 1962). The concept of 'reprogramming factors' was discussed with the *Myod* experiments in mice whereby mouse embryonic fibroblasts were programmed to myoblasts on addition of this cDNA (Davis et al., 1987) and the cloning of Dolly the sheep showed this could occur in mammals (Campbell et al., 1996).

Initial experiments to reprogram mouse and human fibroblasts to induced pluripotent stem cells (iPSCs) occurred using viral overexpression of transcription factors (Park et al., 2008; Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). These methods used retroviral transduction of *Pou5f1*, *Sox2*, *Klf4* and *c-Myc* into mouse embryonic fibroblasts or tail-tip fibroblasts (Takahashi and Yamanaka, 2006). There have been many optimisations to this method to increase the efficiency and to generate 'footprint free' methods, such as using lentiviral vectors, using the cre-lox system to remove the sequences and changing the stoichiometry of the reprogramming factors. While optimisation of these methods continues, it is important to note that these cells may maintain the characteristics of their initial cell type to an extent, and this may be important during differentiation.

While miPSCs are used frequently, they are not discussed in further detail here as they were not used in this project.

1.1.5.3 hiPSCs

Human induced pluripotent stem cells (hiPSCs) are often used in cell culture as they have huge medical potential, especially in cell replacement therapy. These cells are able to overcome a number of issues with hESCs including ethical issues, availability of these cells and host rejection. The similarity of these cells to hESCs as well as to each other is highly debated. It has been shown that different iPS cell lines have a more diverse genetic background, epigenetic signature and differences in reprogramming (Kim et al., 2010; Polo et al., 2010; Rouhani et al., 2014). The differences in reprogramming may be due to the reprogramming process not being complete, as there are many reprogramming practices with differing efficiencies (Osafune et al., 2008). However, since there are still so many differences between the cells referred to as hESCs, it is hard to determine whether the differences seen within hiPSCs are greater and importantly which cell type more closely recapitulates the cell state in the embryo.

The BOBSC-T6/8_b1 (BOBSC) hiPSCs (Yusa et al., 2011) have been used by the Sanger Centre to generate a large library of genetic knockouts using the CRISPR/Cas9 technique, in which there are currently 194 targeted genes (June 2019) (www.sanger.ac.uk/htgt/lims2) (Sanger Centre COMSIG knockout unpublished). Due to the availability of this knockout library, these cells were used throughout this project.

1.1.6 hiPSC and hESC differentiation

1.1.6.1 Differentiation in cell culture

Human ES and iPS cell lines are able to generate all three lineages after gastrulation, as discussed previously, but as more methods are created it is becoming possible to differentiate cells into many differentiated cell types with the possible outcome of forming organs: this includes forming 'mini brains' from cerebral organoids (Lancaster et al., 2013). Protocols for generating the three lineages have improved greatly over time and it is possible to differentiate cells to definitive endoderm with over 90 percent efficiency. This was the protocol used throughout this project. Mapping transcription factor binding dynamics of lineage specification genes has helped to understand the role of DNA methylation in differentiation (Tsankov et al., 2015) whereas studies focussing on histone marks have revealed the importance of modifiers including HDAC1 during gastrulation (Dovey et al., 2010).

1.1.6.2 Definitive endoderm differentiation

Embryonic stem cells can develop into the three lineages: endoderm, mesoderm and neuroectoderm (Figure 2). Definitive endoderm (DE) differentiation has been heavily studied due to its clinical significance in generating whole organs such as the pancreas and liver. Although the protocols allow generation of this cell type with relatively high efficiency and the transcriptomics of this protocol are well understood, a full understanding of the underlying pathways is yet to be elucidated (Chia et al., 2019; Teo et al., 2011, 2012).

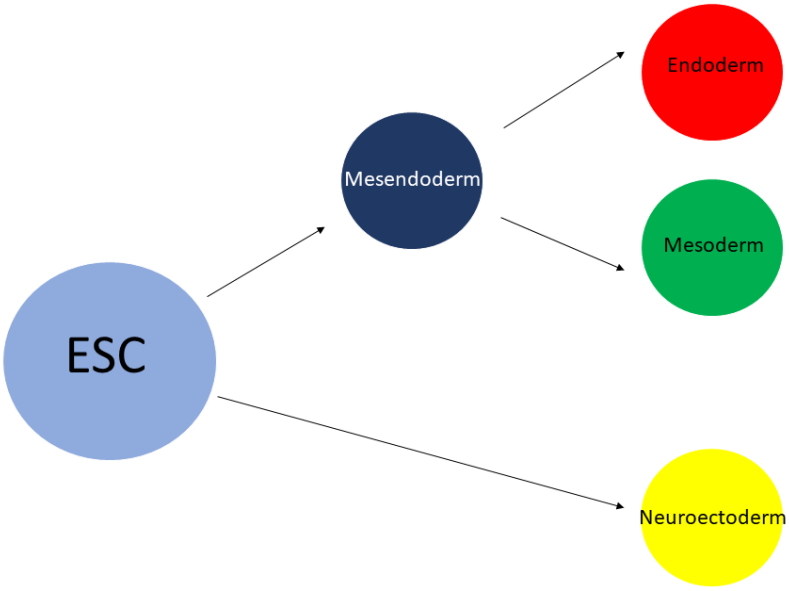


Figure 2. A simplified schematic to show lineage specification in ESCs.

Cells are induced to differentiate using Activin/Nodal signalling, activation of FGF and BMP4 signalling, activation of WNT signalling and inhibition of P13K. This is performed using three recombinant proteins, Activin A, FGF2 and BMP4, and inhibitors of PI3K and GSK3 (Teo et al., 2011). Pluripotent cells differentiate into the primitive streak stage, they then undergo the epithelial mesenchymal transition (EMT) before becoming specified to definitive endoderm. Transition through the EMT causes a dramatic change in cell morphology, visible down a microscope, as well as a switch from the expression of E-cadherin to N-cadherin (Teo et al., 2011). This transition is very important in cancer and has been studied most widely in this context.

In a similar manner to mouse embryonic stem cells, the trio of transcription factors OCT4, SOX2 and NANOG control pluripotency in hES cells. The expression of these genes also regulates the end of pluripotency and the onset of differentiation. When the cells are differentiating, *SOX2* is downregulated within the first 24 hours, *NANOG* by 48 hours and *POU5F1* in 72 hours due to its late downregulation by the definitive endoderm marker SOX17. Two primitive streak (PS) genes known to be expressed in the gastrulating mouse embryo, *EOMES* and *GSC*, have these three transcription factors bound cooperatively to them but, although EOMES is required for human endoderm differentiation, GSC is entirely dispensable (Boyer et al., 2005).

EOMES is expressed from around eight hours in differentiating cells, followed by *MIXL1*, *T* and *GSC* from twelve hours. Therefore, there is a significant amount of overlap between the expression of pluripotency genes and markers of DE. EOMES is thought to be a key cell fate coordinator at this time in differentiation and is its expression is regulated directly by OCT4, SOX2 and NANOG. OCT4 and SOX2 inhibit expression but NANOG upregulates EOMES on the induction of specific differentiation cues. To ensure commitment to differentiation, EOMES is able to downregulate *POU5F1*, *SOX2* and *NANOG* (Figure 3). It is also suggested that BMP4 may play a role in directly inhibiting *POU5F1* and *SOX2* in order to commit to differentiation (Teo et al., 2011).

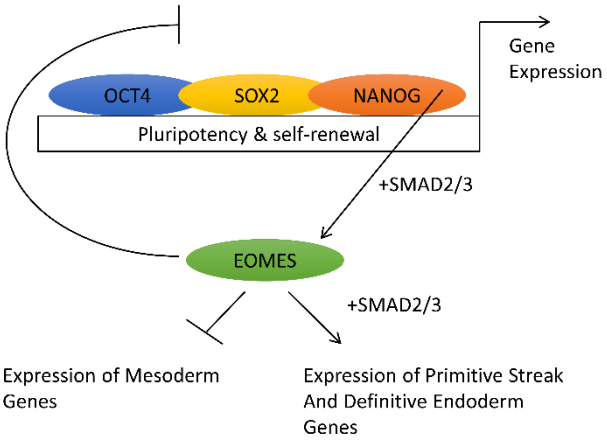


Figure 3. A simplified schematic to show the control of gene expression during differentiation to definitive endoderm from the pluripotent state. EOMES is upregulated, leading to downregulation of pluripotency factors and an upregulation of definitive endoderm specific genes, including SOX17.

EOMES is a key transcription factor thought to push differentiation towards endoderm and prevent the differentiation to mesoderm. This factor has been shown, in complex with SMAD2/3, to bind to a large number of genes involved in differentiation to both the endodermal and mesodermal lineage. Being upregulated early, EOMES is a key protein to activate primitive streak genes including itself, *MIXL1*, *T*, *GSC*, *TBX6*, *FGF8*, *SNAI1*, *SPRY2*, *SPRY4*, *WNT3*, *WNT3A* and *NODAL*. It is also thought to switch on a number of definitive endoderm genes: *SOX17*, *CXCR4*, *LHX1*, *FOXA1*, *FOXA2* and *FOXA3*. It binds to and inhibits a selection of mesoderm specification genes: *MEOX1*, *TBX6*, *KDR*, *FOXC1*, *ISL1*, *PDGFRA*. However, despite EOMES playing a central role at this point in development, its knockdown did not affect the EMT that the cells undergo during differentiation from the primitive streak (Teo et al., 2011).

Specification of definitive endoderm is marked with the co-expression of *SOX17* and *FOXA2*. Differentiation also causes a change in the cell cycle of these cells. They move from having over 60 percent of cells in S phase at one time to the majority of cells being in G1 phase,

indicative of a more differentiated cell type. These features can be validated using flow cytometry to monitor the efficiency of differentiation, discussed in Results I.

However, studies on differentiation have shown that the efficiency of differentiation is not only altered by the culture conditions used to differentiate the cells, but also the confluence of the cells, which is important when comparing differentiation efficiencies of different cell types (Graffmann et al., 2018; Kempf et al., 2016). This is shown to affect the outcome only in the first 24 hours of differentiation, suggesting it is key for lineage specification.

DMSO has also been shown to alter the normal pathway of differentiation, discussed in greater detail in the Results I (Czysz et al., 2015; Pal et al., 2012). This finding is important because DMSO is thought to be able to remove DNA secondary structures, possibly causing the alteration of the differentiation pathway (Kang et al., 2005; Winship, 1989).

1.1.6.3 Mesoderm differentiation

The differentiation to mesoderm is also a key pathway to understand for clinical medicine. The paraxial mesoderm goes on to become the kidney, whereas the lateral plate mesoderm forms heart and blood vessels. Mesoderm differentiation begins in a similar manner to endoderm differentiation due to the shared early cell type, mesendoderm. The protocols for differentiation to mesoderm are varied and relatively complicated, they have not been used at all in this thesis. However, one group reported an efficient protocol for this differentiation using the GSK3i CHIR99021, which was shown to increase mesendoderm gene expression, including T and MIXL and then markers of intermediate mesoderm PAX2 and LHX1 (Lam et al., 2014). The GSK3i addition suggests that this pathway requires a prolonged activation of the WNT pathway for mesoderm differentiation, this is on the contrary to definitive endoderm specification where the pathway is downregulated after 24 hours.

1.1.6.4 Neuroectoderm differentiation

The Studer group has established a method to derive neuroectoderm cells using two small molecule inhibitors, LDN193189 and SB431542, that inhibit the BMP and TGFβ signalling pathways respectively (Tchieu et al., 2017). These two inhibitors are known as dual SMAD inhibitors and cause a cell fate specification to neuroectoderm as marked by the expression of PAX6 and SOX1. The efficiency can be further increased by addition of a WNT pathway activator. Further understanding of this pathway will be incredibly useful in forming neural cells in culture for regenerative medicine.

1.1.6.5 Cell cycle control of differentiation

Numerous experiments have been performed to show that the cell cycle is intrinsically important in controlling pluripotency as well as lineage specification. hES and mES cells have a short G1 phase, thought to be due to the lack of a G1 checkpoint inferred by the resistance to DNA damage (Neganova et al., 2011). This is discussed in greater detail during the DNA damage section. This short G1 phase increases in length throughout differentiation, suggesting that differentiation alters regulation of the cell cycle (Calder et al., 2013).

The propensity of hES cells to differentiate into the different lineages was analysed in Ludovic Vallier's group and it was found that cells are only able to differentiate into certain cell types at certain points in the cell cycle (Pauklin and Vallier, 2013). They see that only cells in early G1 can differentiate towards endoderm, whereas cells commit to neuroectoderm in late G1 phase. They show that this is controlled by Cyclin D which activates CDK4/6 leading to phosphorylation of SMAD proteins. The phosphorylation prevents SMAD movement into the nucleus in late G1 phase, thereby preventing endoderm differentiation and allowing the specification to neuroectoderm. Further work has shown that Cyclin D1 is highly expressed in late G1 where it recruits transcriptional corepressors of endoderm specification genes and coactivators of neuroectoderm genes (Pauklin et al., 2016).

A separate study from the Vallier group has shown that loss of pluripotency, which they name pluripotent state dissolution (PSD), is also dependent on the location of the cell in the cell cycle (Gonzales et al., 2015). They show that S and G2 cell cycle phases act to restrict PSD whereas in G1, when these pathways are absent, PSD could occur. ATM mediated p53 and Cyclin B1 activation can cause S/G2 arrest and this can block the cells in a pluripotent state. This very clearly shows that the activity of cell cycle controlling proteins are able to control cell fate decisions.

1.2 DNA secondary structures

While DNA is often considered to be a stable, double stranded structure holding the genetic information, it is also known to form many different secondary structures *in vivo* (reviewed by Bochman et al., 2012). These structures can cause problems for the DNA polymerase during replication, as well as the RNA polymerase in transcription. There are many non-B-forms of DNA that can form *in vitro* and are thought to occur *in vivo*. These structures include simple hairpins and cruciforms which form in inverted or mirror repeat sequences, triplex DNA that

can form in polypurine DNA stretches including $(CAG)_n$, Z-DNA, R-loops which require RNA and DNA binding, and G-quadruplexes (G4s). The *in vitro* prediction of the locations of these structures often occurs at points of known instability *in vivo* (Burrow et al., 2010). For the purpose of this thesis I will focus on the G4 structure.

1.2.1 The G-quadruplex structure

1.2.1.1 Formation of the G-quadruplex

G4s are able to form in stretches of DNA rich in guanine bases. Four guanines Hoogsteen base pair to form a G-quartet (Gellert et al., 1962), stabilised by a central monovalent cation such as K^+ or Na^+. These G-quartets stack through π-π interactions to form the G-quadruplex (Sundquist and Klug, 1989; Williamson et al., 1989). These G4s can be tetramolecular, bimolecular or unimolecular, formed using a single strand of DNA, as is most significant in the context of DNA replication (Figure 4). The G4 structure formation occurs preferentially on single stranded DNA, such as during transcription and replication.

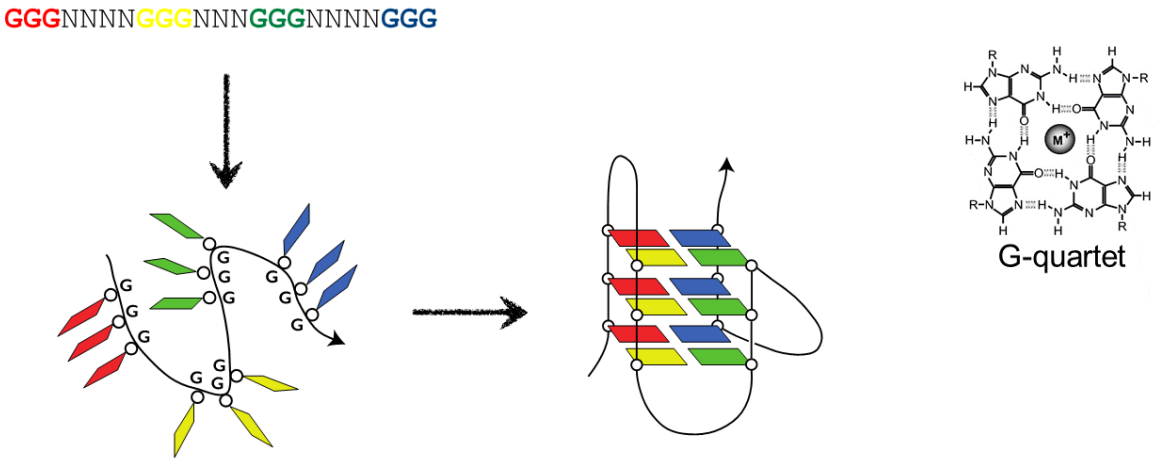


Figure 4. The canonical G4 requires four runs of three guanines on one strand of DNA to fold into a unimolecular G4. Each planar stack is formed of a G-quartet containing four guanines and a central metal ion to stabilise the structure. Adapted from Lerner et al., 2019; Šviković and Sale, 2017.

1.2.1.2 Biological significance of G4s

The role, if any, that these structures played *in vivo* was and still is heavily debated (Maizels and Gray, 2013; Murat and Balasubramanian, 2014; Rhodes and Lipps, 2015; Tarsounas and Tijsterman, 2013). However, the initial evidence for the significance of these structures was shown when G-rich sequences, seen in telomeres, gene promoters and immunoglobulin switch regions, could form G4s and they proposed that self-recognition of these motifs could attach the four chromatids during meiosis (Sen and Gilbert, 1988). These structures have since also been shown to be enriched in ribsomal DNA, transcription start sites and mitotic and meiotic

DSB sites (Capra et al., 2010; Eddy and Maizels, 2006; Hershman et al., 2008; Huppert and Balasubramanian, 2005; Nakken et al., 2009). These sequences have also been found to be conserved between yeast and human genomes (Capra et al., 2010). Since these structures are enriched at key locations in the genome it has been suggested that they play a number of biological roles in these cells.

Telomeric repeats have high numbers of G4 structures due to the repeat sequence (AGGGTT) acting as a canonical G4 (Huppert and Balasubramanian, 2005). Therefore G4s are predicted to form in the telomeres of many species. *In vitro* studies have shown that the telomere associated proteins, TEBPα and β in ciliates and Rap1 in *Saccharomyces cerevisiae*, promote G4 formation (Fang and Cech, 1993; Giraldo and Rhodes, 1994; Paeschke et al., 2005; Sundquist and Klug, 1989). These structures have also been linked to specification of DNA replication origins in vertebrates (Besnard et al., 2012; Cayrou et al., 2012; Valton et al., 2014).
G4s are also thought to play a role in the regulation of transcription through a number of means. For example, the *c-MYC* locus is commonly used in the discussion of G4s as it has been highly studied. It is of added interest due to its role in cancer. This locus contains the nuclease hypersenstive element III_1 downstream of the *MYC* promoter which controls the majority of transcription from this locus. In this element there is a G4 motif that has been shown to form a G4 *in vitro* (Simonsson et al., 1998). Mutagenesis of this region to prevent the formation of a G4 show that this is no longer able to repress transcription of this region (Siddiqui-Jain et al., 2002). Therefore, in this context, the G4 is shown to act as a transcriptional repressor. Numerous tumour supressor and proto-oncogenes besides *MYC* have been shown to contain DNA G4s, as well as RNA G4s, discussed later, and this may be key for their regulation (Welsh et al., 2013).

The myosin D (MyoD) protein family of transcription factors has been highly studied for its role in regulating transcription through G4s. These proteins bind to the promoters of muscle-specific genes to regulate muscle development. *In vitro* studies have shown that MYOD homodimers bind to G4s enriched in muscle specific genes. However, MYOD-MYOE heterodimers do not bind to these G4s and instead bind to an E-box in the promoters of these genes and upregulate transcription (Shklover et al., 2010; Yafe et al., 2008).

The *HRAS* promoter had been shown to contain two regions of DNA which are each able to fold into a G4 (Cogoi et al., 2014). Importantly, these have been shown to be bound to two zinc-finger transcription factors MAZ and Sp1. The authors have suggested that the MAZ protein is

able to bind and then unwind these G4 structures. This knowledge was used to generate anticancer molecules used to prevent transcription of *HRAS* (Cogoi et al., 2014).

The evolutionary significance of G4 forming sequences is debated due to the number of organisms with A/T-rich genomes, including *Plasmodium falciparum* and *Dictyostelium discoideum* (Szafranski et al., 2005). While direct evidence of the biological role of G4s *in vivo* is not vast, they are known to form potent replication blocks and this has great impacts for gene expression which is discussed next. Regardless of their role, these structures are able to form and are therefore able to cause problems during replication and transcription.

1.2.1.3 G4 prediction tools

Estimations suggest that there are around 700,000 sequences in the human genome that can form G4s but this estimation is highly dependent on the stringency of the algorithm, as many G4s that exist do not conform to the cannonical sequence ((Huppert and Balasubramanian, 2005; Sahakyan et al., 2017; Todd et al., 2005).

The 'Quadparser' algorithm was developed by Shankar Balasubramanian's group and simply allowed a search of DNA sequences of the nature $(G_{\geq 3}N_{1\text{-}7})_4$ (Huppert and Balasubramanian, 2005). This motif was predicted to form in physiological conditions and was therefore suggested to be a good initial predicter (Hazel et al., 2004). However, the canonical $(G_{\geq 3}N_{1\text{-}7})_4$ sequence has been long known to not include a number of G4 structures shown to form (Patel and Hosur, 1999). Biophysical studies have also shown large differences in the thermal stability between different G4s (Hazel et al., 2004).

Another predictive tool, quadruplex-forming G-rich sequence (QGRS) mapper, was generated using a looser G4 definition: $(G_{\geq 2}N_n)_4$ (D'Antonio and Bagga., 2004). This allows you to input a sequence of DNA and outputs a list of potential G4s ranked with a G-score. This G-score informs the user of the likelihood of a structure forming in the DNA based on the number of G-quartets stacked, the length of loops and the symmetry of the structure.

Eddy and Maizels used a 'sliding window' technique to calculate the potential of a G4 forming (Eddy and Maizels, 2006). This strategy calculated the likelihood of G4 formation based on the density of guanines within a region of the DNA; the score received is independent of the length of the sequence.

While giving a quick look at the possible G4 potential, these methods were not the most reliable as the prediction tools were extrapolated from a small data set. Some of the more recent predication methods are likely to give more biologically relevant answers. These allow combinations of longer loops, mismatches, bulges, location of cytosines and structural features: these included Pgsfinder, Quadron and G4-hunter (Beaudoin et al., 2014; Bedrat et al., 2016; Hon et al., 2017; Mukundan and Phan, 2013; Sahakyan et al., 2017). These approaches have been tested on much larger datasets and have improved predictions, especially of non-canonical G4s. Furthermore, groups have attempted to create lists of genes containing G4s (Zhang et al., 2008).

1.2.1.4 Evidence that G4s form *in vivo*

The G4 structure has been visualised using antibodies suggested to be specific to G4s, although the specificity of this is debated. Visualisation of DNA G4s in the ciliate *Stylonychia lemmae* was achieved using antibodies specific to the telomeric repeat of this organism (Schaffitzel et al., 2001). This provided early evidence that these structures existed in the cell. The BG4 antibody was generated using phage display and used to visualise G4s in DNA using immunofluoresence (Biffi et al., 2013). This study also showed an increase in these structures after the addition of a G4-binding ligand, suggesting that they do stabilise these structures *in vivo* and that the antibody shows some degree of specificity. However, the monoclonal antibody 1H6 which was initially thought to recognise G4s was shown to cross-react with thymidine-rich single stranded DNA, highlighting the problems with using these techniques (Kazemier et al., 2017).

1.2.1.5 G4 prevelance in the mitochondrial genome

Most of the work on G4s has been done using nuclear DNA, however it is becoming clear that mitochondrial DNA (mtDNA) also contains these structures. Around 90 potential locations for G4 formation have been found in the mtDNA . These are biased to the heavy strand due to the high guanine content (Dong et al., 2014; Zybaĭlov et al., 2013). It has also been shown that a G4 is involved in controlling the switch between transcription and replication (Agaronyan et al., 2015). Specific ligands have also been developed to target the mitochondrial genome (Huang et al., 2015). These G4s are not the focus of this thesis, but it is worth noting as many perturbations generated by G4 ligands are likely to cause problems in the mitochondria too.

1.2.1.6 RNA G-quadruplexes

While much of the initial work on G4s focussed on the DNA G4s, there has been subsequent analysis of RNA G4s which are also thought to have significant biological impact. These

structures were thought to have similar propensities to form, and similar properties. However, the chemical differences between DNA and RNA have revealed changes by using biophysical measurements. RNA has an extra 2'-OH group which was shown to allow additional intramolecular interactions in the loop regions and therefore enhances the stability of these structures compared to DNA G4s (Zhang et al., 2010). RNA G4s also prefer parallel G4 folding, compared to DNA which is able to fold into parallel, antiparallel and mixed structures (Zhang et al., 2010). However, it is also important to remember that RNA is much more likely to form a number of other secondary structures, compared to DNA, due to its single stranded nature. These considerations are likely to alter the probability of RNA folding into a G4.

As for DNA G4s, RNA G4s are thought to play a key role in a number of processes (Bugaut and Balasubramanian, 2012; Fay et al., 2017). One such role is in translation whereby a G4 in the 5' or 3' UTR, as well as in the coding sequence, has been seen to prevent efficient translation (Arora and Suess, 2011; Crenshaw et al., 2015; Endoh et al., 2013; Kumari et al., 2007). These G4s have also been shown to alter splicing codon choice (Burley et al., 2017; Marcel et al.), RNA localisation (Subramanian et al., 2011) and act as the termination structure in ρ-independent mitochondrial transcription termination (Wanrooij et al., 2010). RNA G4s also have a stronger binding affinity to Polycomb repressive complex 2 (PRC2) than an unstructured G-rich motif or duplex RNA, suggesting a role for these structures in epigenetic control (Wang et al., 2017). These RNA G4s have also been visualised in cells using BG4, as discussed previously, using a ligand to stabilise the structure (Biffi et al., 2014).

Non-coding RNA G4s have also been shown to form in tRNAs encoded in the nuclear and mitochondrial genome. Some work has shown that the G4 structure is required for the degradation of these non-coding structures produced in the mitochondria (Pietras et al., 2018). Another group has shown that cytoplasmic tRNAs are cleaved during the stress response to produce fragments, and these assemble into intermolecular G4s which are suggested to play a further role in the stress response (Lyons et al., 2017).

While the aims set out in this thesis are to understand how replication impediments alter differentiation in a DNA replication context, the ability of RNA to form G4s is likely to impact the experiments and should be considered when examining the results.

1.2.2 G4-processing enzymes

The existence of G4 structures poses the problem of replicating these stretches of DNA by replicative polymerases (Woodford et al., 1994). Interestingly, the unwinding of the DNA duplex lends itself to the formation of G4s on both strands but not equally. Replication of the leading strand of DNA is known to be continuous whereas lagging strand synthesis occurs via formation of Okasaki fragments and occurs discontinuously. The nature of lagging strand replication was thought to lend itself to G4 formation due to the increased length of time of it being single-stranded (Cheung et al., 2002; Ding et al., 2004). However, the lagging strand is thought to be coated in RPA which prevents the formation of these G4s (Safa et al., 2016). Evidence now suggests that G4s are able to form on the leading strand during DNA replication, shown in the BU1 assay discussed later (Schiavone et al., 2016).

This means that the process of DNA replication can enhance the formation of G4s on the DNA, and this can also occur during transcription. Increasing the length of time of the DNA being single stranded, for example by slowing replication, can also increase the chance of a possible G4 forming.

Certain helicases, primase and polymerases have been implicated in the replication of the G4. *In vitro* the majority of human helicases tested are able to unwind G4 structures, and those that do are often associated with a disease phenotype such as the RecQ helicases WRN and BLM, and FANCJ (Huber et al., 2002; London et al., 2008; Mohaghegh et al., 2001; Ribeyre et al., 2009; Sanders, 2010). However, these disease phenotypes are not clearly linked to G4-processing defects. Mutations in the human FANCJ protein are associated with Fanconi anaemia and these patients show a phenotype suggesting a lack of ability to replicate G-rich regions of the DNA: their DNA shows deletions of these G-rich regions (London et al., 2008). In a similar manner to the FANCJ mutations in human patients, mutations in the *Caenorhabditis elegans* DOG-1 helicase, a relation of FANCJ, can cause large scale deletions in G-rich sequences (Cheung et al., 2002; Kruisselbrink et al., 2008). These types of studies highlight the importance of proper processing of these G4 structures, and the requirement of specific helicases in this process.

1.2.3 Maintenance of epigenetic marks through DNA replication

While it is vitally important for the cell to perfectly replicate its DNA across generations in order to maintain fidelity, it is also important that epigenetic marks transferred such that gene expression state is not lost. The structure of DNA suggested a simple, elegant model as to its

replication (Watson and Crick, 1953), but replication of marked nucleosomes on the DNA did not lend itself to a straightforward explanation. As discussed in the earlier sections, histone modifications and DNA methylation, both forms of epigenetic information, play key roles in embryonic development. During replication the histone marks associated with a genomic region must be returned to this region and this has been shown to be correct using sequencing methods (Reverón-Gómez et al., 2018). The replicative helicase removes the histones from the DNA and these are then randomly distributed to each of the leading and lagging DNA strands by the histone chaperones ASF1 and CAF1 (De Koning et al., 2007). In order to maintain the number of histones on each DNA strand, new, unmodified histones are added with these recycled histones and modifications are transferred from the old to new histones to maintain the epigenetic information (Bannister et al., 2001; Hansen et al., 2008; Lachner et al., 2001; Margueron and Reinberg, 2010). This process must be performed in a timely fashion to maintain the coordination between DNA synthesis and the histone marks associated with these regions.

Much of the published work looking at the replication of G4s and its effect on epigentic instability has been done in the DT40 cell line, a chicken B cell line derived from leukosis virus induced bursal lymphoma (Baba et al., 1985), initally used for its ease of genetic manipulation (Buerstedde and Takeda, 1991). This cell line has an easily observable phenotype which arises as a result of a G4 sequence 3.5kb from the transcription start site (TSS) of the BU-1 locus (Sarkies et al., 2012).

The BU1 gene encodes a cell surface receptor that is uniformly highly expressed in wildtype DT40 cells and can be monitored using flow cytometry. However, clonal expansion of these cells shows expression of the BU-1 locus decreases in a stochastic manner at every replication cycle in cells either stabilised with G4-binding ligands, containing genetic knockouts of processing enzymes or experiencing nucleotide pool depletion. The general proposed mechanism for this is a blockage of the DNA replication fork at the G4 secondary structure and restart downstream. This causes a loss of the histone modifications associated with this tract of DNA as new, unmarked histones are deposited (Figure 5). Monitoring of this phenomenon is discussed in the next section.

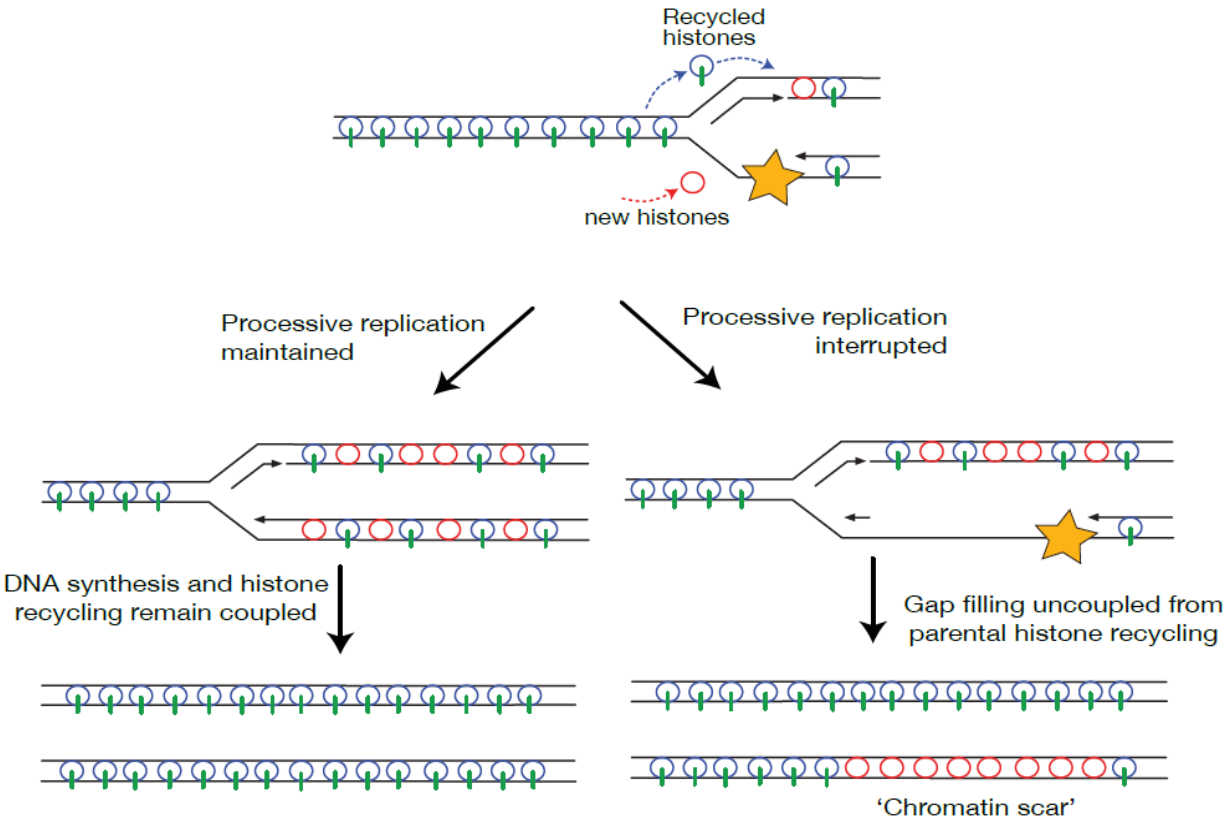


Figure 5. When DNA replication is uncoupled from the deposition of the original histones, the chromatin structure can change and alter gene expression. This uncoupling can be caused by a number of factors including DNA damage, secondary structures and replication stress, shown with the yellow star, in this instance a G4. Adapted from Sarkies et al., 2010.

1.2.3.1 Using the BU1 assay to detect G4-dependent epigenetic instability

Epigenetic instability is defined here as a loss of chromatin marks in a region of DNA, often resulting in changes in gene expression. This has been studied in our group in great detail, initially at the ρ-globin locus and then the BU-1 locus of DT40 cells (Sarkies et al., 2010, 2012; Schiavone et al., 2014).

Epigenetic instability occurs due to uncoupling that is thought to occur because the replicative DNA helicase moves through the G4 structure whereas the polymerase gets stuck. This has been suggested to leave a ssDNA gap where the G4 forms. This replication process leaves gaps in the leading strand of DNA. The replication of these gaps occurs later, meaning that parental histones can no longer be transferred to the daughter DNA strand and naïve, unmarked histones are deposited such that the original histone markings are lost, which leads to changes in gene expression (Figure 5). While the BU-1 locus has proved to be a good system to analyse the role of G4s in inducing epigenetic instability, this phenomen is capable of happening at any location where the helicase and polymerase are uncoupled. This can include a slowed rate of DNA

synthesis (Papadopoulou et al., 2015), described later using low dose hydroxyurea, DNA damage, genetic knockout of factors involved in processing of secondary structures and stabilising secondary structures on the DNA, as discussed using G4-binding ligands. Using these model systems, G4-processing enzymes have been implicated in processing genomic G4s, acting in concert at both the 5' and 3' ends of the structure.

REV1 is a Y family DNA polymerase known to play a role in translesion synthesis occuring after DNA damage, in order to maintain replication fork progression (Edmunds et al., 2008; Jansen et al., 2009). In cells lacking *REV1*, postreplicative gap filling occurs in gaps of 400-3000 base pairs, around 20 kilobases behind the replication fork (Lehmann, 1972; Lopes et al., 2006). This distance allows an uncoupling of genetic and epigenetic marks at specific DNA secondary structures to be seen in these REV1 deficient cells. This is due to the gap filling being uncoupled from bulk DNA replication and therefore the loss of histones associated with this region as they are displaced when the replicative helicase unwinds the DNA. At the β-globin locus in DT40 cells this is seen through the derepression of ρ-globin (Sarkies et al., 2010). This work showed that the REV1 polymerase played a role in the processing of secondary structures, namely G4s, a previously unexpected role.

Further work has focussed on the BU-1 locus in DT40, introduced in Section 1.2.1.3, as this offers a powerful system to enable the study of epigenetic instability. The BU-1 locus contains a putative G4 between exons two and three on the leading strand, around 3.5 kilobases downstream of the TSS (Figure 6) (Sarkies et al., 2012). The BU1 protein is ubiquitously expressed on the surface of DT40 cells and can be monitored using permeabilised flow cytometry. When there is an uncoupling of DNA synthesis and deposition of recycled histones, as seen in *REV1*$^{-/-}$ cells, the epigenetic information is lost and expression of BU1 is lost in a stochastic manner at each replication cycle, this is associated with a loss of H3K4me3 and H3K9/12ac (Sarkies et al., 2012; Schiavone et al., 2014). Using this elegant assay, FANCJ, WRN/BLM (in a redundant manner) and the primase polymerase PRIMPOL have been shown to also play a role in G4-processing during DNA replication (Sarkies et al., 2012; Schiavone et al., 2016). FANCJ and WRN and BLM helicases and REV1, PrimPol, Pol κ and Pol η polymerases have been shown to play a role in G4-processing in a number of other studies (Bétous et al., 2009; Crabbe et al., 2004; Huber et al., 2006; Rey et al., 2009; Wu et al., 2008). Interestingly, there is some evidence to suggest that these polymerases have preference for the type of G4 they replicate such that REV1 is more able to process G4s containing longer loops,

whereas PRIMPOL is involved in priming after shorter G4s with a higher melting temperature (Schiavone et al., 2014, 2016).

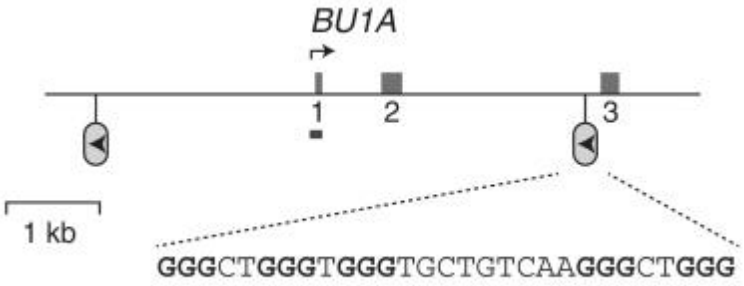


Figure 6. The structure of the BU-1 locus in DT40, taken from Sarkies et al., 2012. The putative G4 is in the second intron, towards exon three, with the sequence shown.

1.2.4 G4-stabilising ligands

The prospect of targeting these G4 structures has led to a significant focus on the development of ligands to bind and stabilise these structures in the DNA and RNA: these are referred to as G4-binding ligands. An initial reason for targeting was to block the action of telomerase on telomere elongation, for example in cancer cells, by using a G4-binding ligand to stabilise the structure in these repeats (Mergny et al., 2002; Neidle and Parkinson, 2002; Oganesian and Bryan, 2007; Patel et al., 2007).

While there are a vast range of G4-binding ligands which bind to subsets of these structures with varying affinities, the features of these small molecules are similar. As mentioned previously, G4s have a large π-surface therefore, in order to increase the π-π interactions, most of the small molecules that bind to G4s also have large π-systems. The ligands also tend to be positively charged in order to bind to the negatively charged DNA, but this is non-specific. Due to this lack of specificity many ligands will also bind to DNA not containing G4s in a less specific manner, hence causing problems in cell systems where there are much greater quantites of duplex DNA compared to G4s. Ligands are thought to be able to bind to G4s through tetrad-stacking or binding to loop or groove regions, suggesting further specificity that could be used for medical targeting of G4s (Kimura et al., 2007; Le et al., 2015). In order to measure the binding of ligands to G4s a number of biophyscial techniques have been used including surface plasmon resonance (SPR) (Redman, 2007).

A vast array of G4-binding ligands have been used, both from natural sources, and those rationally generated to bind and stabilise G4s. These ligands are closely related to one another: TMPyP4, Telomestatin, Pyridistatin (PDS), PhenDC3, N-Methyl Mesoporphyrin IX (NMM),

and PIPER (Figure 7). In this thesis NMM, PDS and PhenDC3 were used to stabilise G4 structures.

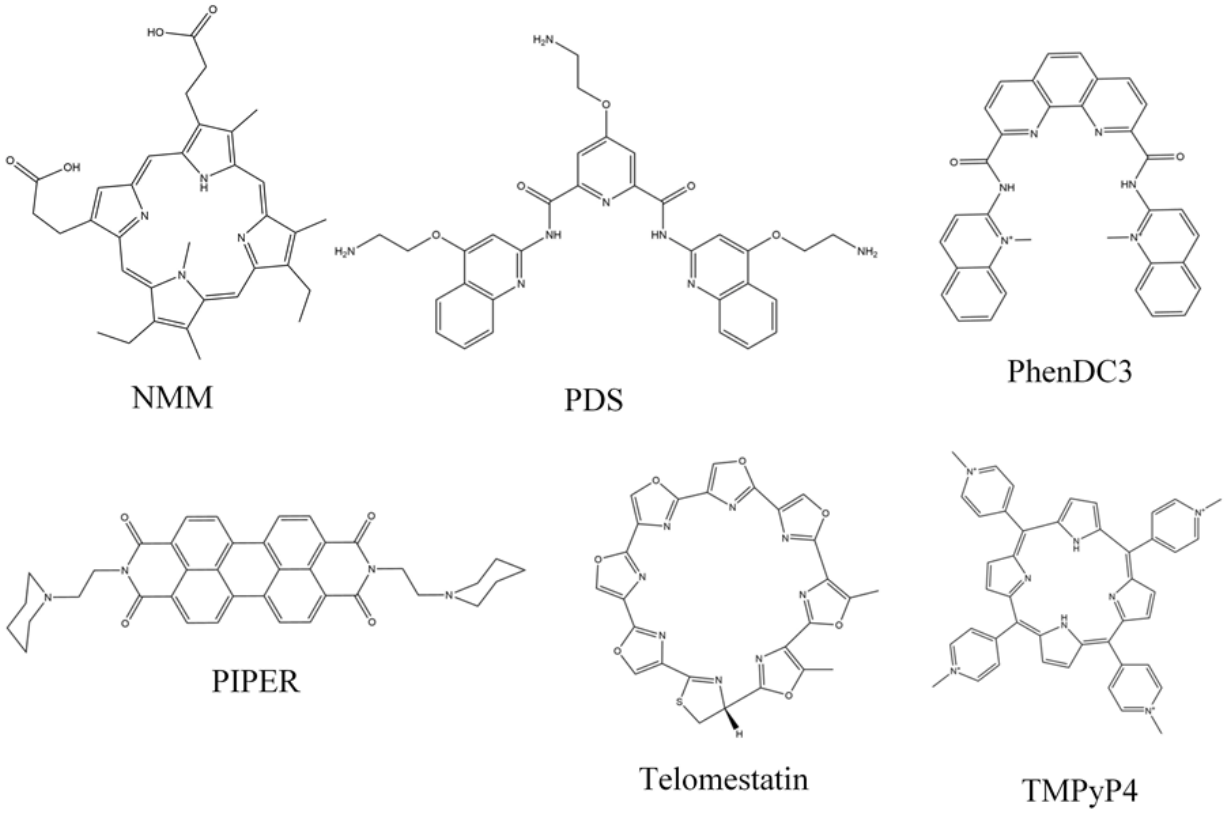


Figure 7. Structures of six commonly used G4-binding ligands. The structures share large delocalised regions in aromatic rings and mostly sit on top of the G-quartet stacks to increase the π-π bonding.

NMM was initially developed as a DNA aptamer and was later found to be selective for G4s over ssDNA, dsDNA, dsRNA, RNA-DNA hybrids, Z-DNA, triplex and Holliday junctions (Li et al., 1996; Nicoludis et al., 2012). A later developed drug, PDS, was rationally designed based on previously known G4-binding ligands: it contains a planar electron system but is more flexible to make it suited to binding more structures. PDS was shown to stabilise telomeric DNA and alter the shelterin complex, thus inducing a DDR (Rodriguez et al., 2008). It has also been shown to cause transcription and replication-dependent damage by binding specifically to G4s as measured by γH2AX foci (Rodriguez et al., 2012). PhenDC3 is a member of the bisquinolinium family that has been shown to bind G4s with high selectivity and specificity by overlapping the G-quartet in a crescent shape (De Cian et al., 2007; Moruno-Manchon et al., 2017; Rigo et al., 2016). PhenDC3 has also not been linked to DNA damage response like PDS.

Each of these three drugs was shown to induce BU1 instability in the DT40 system and was therefore appropriate for trial in this thesis, with varying success as discussed in the results chapters. None of these drugs have the perfect properties for use in cell systems due to size, toxicity and specificity and a new drug was developed in an attempt to combat these issues (Guilbaud et al., 2017). While this is also discussed in the results chapters, it was not appropriate to use in either human or mouse system.

1.2.5 Hydroxyurea induced nucleotide pool depletion

Replication stress induced by nucleotide pool imbalance, for example using the addition of hydroxyurea (HU), can lead to a decrease in the DNA polymerase speed leading to a local uncoupling of the polymerase and DNA helicase (Byun et al., 2005; Pacek and Walter, 2004; Papadopoulou et al., 2015), as mentioned in the BU1 assay. HU induces ribonucleotide reductase (RNR)-inhibition and a depletion of dNTPs. This uncoupling leads to formation of regions of ssDNA and usually activation of DDR checkpoints (Byun et al., 2005; Pacek and Walter, 2004). Using the BU1 assay, it was shown that chronic, low doses of HU gave rise to stochastic loss of expression of BU1. First the loss of the active chromatin marks H3K4me3 and H3K9/14ac and then the appearance of the repressive mark H3K9me3 and the DNA damage marker γH2A.X (Papadopoulou et al., 2015).

1.3 The DNA damage response

Replication impediments cannot be discussed without focus on the DNA damage response (DDR), and its importance in all cells including during differentiation. Thousands of lesions are formed in the DNA in every human cell each day as a result of endogenous and exogenous damage as well as errors during DNA replication (Lindahl and Barnes, 2000), as discussed in reference to G4s. These structures must be repaired to enable processivity of the replication and transcription machinery, as well as to allow faithful transmission of genetic information to the next generation. Cancer is a fundamental example of the result of this process going wrong. There are many forms of environmental DNA damage humans expose themselves to including ultraviolet (UV) light, being commonly known to cause skin cancer in humans, and tobacco, which can cause lung cancer.

In order to deal with these various forms of DNA damage, cells have evolved the DDR whereby cells are able to detect damage, signal that they are damaged and repair the damage to the greatest extent possible at the time. DNA damage arises in the DNA in a range of forms due to

the range of DNA damaging agents, and the type of damage necessitates the type of repair required.

1.3.1 The DDR in somatic cells

1.3.1.1 DNA repair pathways

There are three pathways that can be used to repair DNA damage on one strand of the DNA. Base-excision repair (BER) allows the removal of a non-helix distorting, damaged base from the DNA throughout the cell cycle. The base is recognised by a DNA glycosylase that removes the base, leaving an AP site which is cleaved by an AP endonuclease. The single stranded gap can be filled in using long (two to ten nucleotides) or short patch (single nucleotide) repair with a polymerase and ligase (reviewed by David et al., 2007; Hoeijmakers, 2001). On the other hand, nucleotide excision repair (NER) is able to recognise helix-distorting lesions and acts via two pathways, transcription-coupled NER and global-genome NER (reviewed by Hoeijmakers, 2001). These pathways act in a similar manner to BER but require the excision a larger region of DNA, 22-30 nucleotides. Mismatch repair (MMR) detects mismatches of bases in the DNA as well as insertions and deletions and repairs the bases (reviewed by Jiricny, 2006).

When a double strand break ensues, there are two key mechanisms controlling the repair processes: non-homologous end-joining (NHEJ) and homologous recombination (HR) (reviewed by Lieber, 2008; San Filippo et al., 2008). These two strategies compensate for one another but offer the cell differing levels of accuracy and therefore are employed to different levels in each cell type. The phase of the cell cycle also controls the choice of pathway in the cell. Due to the higher accuracy of HR, this is most frequently used in stem cells in order to maintain fidelity.

In standard NHEJ, the Ku70/Ku80 heterodimer recognises DSBs in the DNA and activates DNA-PKcs. When classic NHEJ is not active there are two other pathways that can be used: microhomology-mediated end-joining (MMEJ) and alternative end-joining (AEJ) but these are always error-prone (McVey and Lee, 2008). These pathways are active throughout the cell cycle as they do not require a homologous DNA template. This is on the contrary to HR, which requires a sister-chromatid and therefore can only occur in S and G2 phases of the cell cycle. HR starts with the creation of ssDNA using the MRN complex, and then the DNA invades the undamaged template using RAD51, BRCA1 and BRCA2 (San Filippo et al., 2008). HR can also be used to repair interstrand crosslinks with the help of the Fanconi Anaemia pathway (Kennedy and D'Andrea, 2005).

1.3.1.2 Types of DNA damage

While there are many forms of DNA damage that cause distinct lesions in the DNA and therefore recruit different repair factors and cause the activation of different pathways, the main forms of DNA damage dealt with in this thesis are those induced with G4-binding ligands, UV irradiation, methyl methanesulfonate (MMS) and hydroxyurea. As discussed previously, some G4 ligands have been shown to induce DNA damage in the form of DSBs, possibly at the location of the G4s they bind (Rodriguez et al., 2012), but this is not thought to be a general feature of these ligands.

Ultraviolet damage induces two mutagenic, cytotoxic lesions in the DNA: cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs) (Yagura et al., 2011). Most UV damage is repaired using the NER mechanism described above. Two human diseases with mutations in the NER pathway, Xeroderma pigmentosum (XP) and Cockayne's syndrome (CS), display increased sensitivity to UV light (Cockayne, 1946; Hebra F, Fagge CH, 1874).

MMS is an alkylating agent which modifies guanine, adenine and cytosine to 7-methylguanine, 1-methyladenine, 3-methyladenine and 3-methylcytosine respectively. The modified guanine causes base mispairing whereas adenine induces replication blocks (Beranek, 1990). The damage induced by MMS is repaired using the BER pathway, described above, and the action of DNA alkyltransferases (Lindahl and Wood, 1999).

HU was also used in a manner to induce chronic replication stress but is also able to induce DNA damage in cells. HU acts as an inhibitor of RNR, thereby exerting its effects on nucleotide pool depletion (Shao et al., 2004). However, it is now thought to induce damage including base oxidation and depurination (Sakano et al., 2001). These are important considerations when using the drug in biological systems, especially as it is used in the treatment of many diseases and as a chemotherapeutic agent.

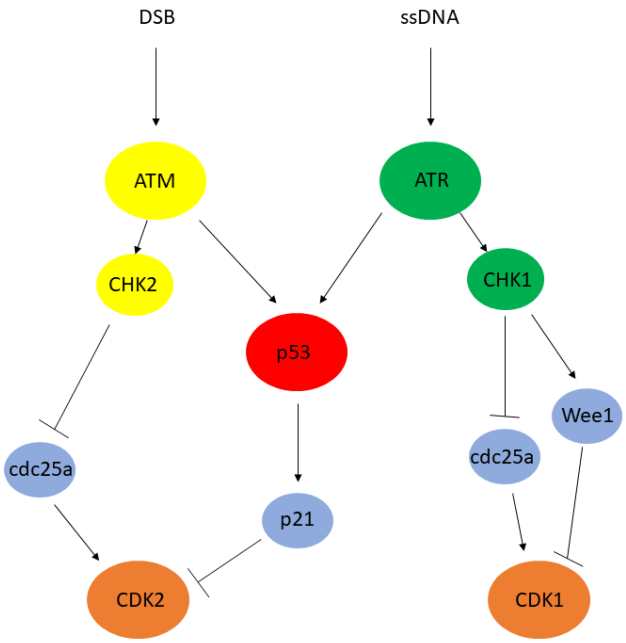
1.3.2 Checkpoint activation after DNA damage

In order for the cell to process and deal with DNA damage, it must activate the DNA damage checkpoint in order to induce a DDR and repair the damage. It is essential for the cell to control this checkpoint to maintain the genetic identity of the cell, and to switch it off when the repair has occurred so as to allow the cell to re-enter the cell cycle. These control pathways are incredibly important as seen by the number of human diseases associated with mutations in DDR genes (Ciccia and Elledge, 2010). This pathway acts to reduce CDK activity in order to

slow down or inhibit cell cycle progression at the G1/S, S and G2/M checkpoints in the cell cycle. It also regulates DNA repair, and when the level of damage is too high, senescence or apoptosis (Kruse and Gu, 2009). The DDR is also suggested to induce cellular differentiation as a mechanism to protect the organism, probably to protect from the cancer stem cell phenotype.

Three members of the phosphatidylinositol-3-kinase-like kinase family control these signalling pathways, ATM, ATR and DNA-PK, and they phosphorylate vast numbers of proteins to induce a response (Matsuoka et al., 2007). These proteins act in concert with the checkpoint protein kinases that they activate, CHK1, CHK2 and MK2 (Liu et al., 2000; Matsuoka et al., 1998; Reinhardt et al., 2007) (Figure 8).

Figure 8. The canonical roles of ATM and ATR kinases in the DNA damage response. Double strand breaks activate ATM causing its phosphorylation, it is subsequently able to phosphorylate and activate a number of downstream proteins causing the cell cycle to be halted.



ATM is recruited to DSBs whereas ATR acts on a larger range of damage that cause ssDNA coated with the single stranded DNA binding protein RPA (Zou and Elledge, 2003). ATM is recruited to DSBs by the MRN complex (Lee and Paull, 2005; Uziel et al., 2003) and phosphorylates a range of proteins, including H2A.X, CHK2, BRCA1 and p53 in order to induce a response (Banin et al., 1998; Canman et al., 1998). The phosphorylation of H2A.X occurs minutes after damage and rapidly spreads over hundreds of kilobases of DNA, this helps to recruit DNA repair proteins and chromatin remodellers (Meier et al., 2007; Savic et al., 2009). Both ATR and ATM appear to be required for DSB repair, suggesting they have non-redundant functions (Brown and Baltimore, 2003; Wang et al., 2004).

ATM is not an essential protein in mice or humans but ATR is, as shown by embryonic lethality in mice and attempted knockouts in human cell lines (Brown and Baltimore, 2000, 2003; Cortez et al., 2001; de Klein et al., 2000). It is suggested that ATR may protect the genome constantly during DNA synthesis through the regulation of nucleotide levels, origin firing, fork progression and control of movement through the cell cycle (Byun et al., 2005; Lopes et al., 2006). ATM and ATR kinases show significant crosstalk in the DDR, for example ATR is also able to phosphorylate H2A.X and recruit ATM (Ward and Chen, 2001), and ATM is able to be phosphorylated by ATR directly, inducing its activation (Stiff et al., 2006).

1.3.3 The DNA damage response in differentiation

The work in this thesis examines the role of the DDR in pluripotent cells and during differentiation. While this process has not been studied exhaustively in this context, there is literature suggesting that proteins regulating the DDR may play a role in normal differentiation (Fujita et al., 2008; Oka et al., 2019; Sherman et al., 2011; Weiss and Ito, 2015). There is also significant evidence that stem cells process damage in a different way to somatic cells (Cervantes et al., 2002)(reviewed by Vitale et al., 2017), as would be expected from the nature of this cell type. Maintaining cells in an undifferentiated state after DNA damage has occurred is often preferred due to the higher level of DNA damage repair factors and increased homologous recombination (Momcilovic et al., 2010; Tichy et al., 2010). The importance of DDR factors during differentiation is discussed below.

1.3.3.1 Histone variants required during differentiation

There are a number of non-canonical histone variants that are used only in certain cases in the cell. The phosphorylation of histone variant H2A.X has been implicated in stem cell self-renewal in mESCs and miPSCs, however it is not clear whether this is also the case in human cells (Turinetto et al., 2012). These γH2A.X foci did not correlate with any DDR factors such as 53BP1 and RPA, and decreased on differentiation (Banáth et al., 2009). Previous groups have highlighted H2A.X during the EMT (Singh et al., 2015; Weyemi et al., 2016a). Singh et al describe H2A.X phosphorylation, by ATM kinase on the canonical serine-139, and how this is required to induce HMGA2 transcription. H2A.X has the interesting property that its phosphorylation acts to destabilise the nucleosome and impair the H1 histone from binding (Li et al., 2010). This is important for the binding of HMGA proteins described in this paper as this family competes with H1 for binding to linker DNA, also loosening the chromatin (Catez et al., 2004; Kishi et al., 2012). In a separate study, Weyemi et al show that loss of H2A.X in the MCF10A and HCT116 cell line activates the EMT program (Weyemi et al., 2016a, 2016b).

Taking these two results together suggests that unphosphorylated H2A.X maintains the cells in an epithelial state. While H2A.X mice are viable, they display a number of abnormalities including growth retardation and immune deficiency, as well as male mice being infertile (Celeste et al., 2002).

The proposed phosphorylation of histone H2A.X at the EMT is associated with mass transcriptional activation caused by transcription factors binding and mediating this transition (Singh et al., 2015). However, recent data from the Rosenberg group focusing on *Escherichia coli* has shown that transcription factor binding is a sufficient block to induce replication fork stalling and reversal, leading to DNA damage at these transcription factor binding sites (Xia et al., 2019). The tight interplay between H2A.X controlled gene expression and DNA damage makes it hard to decipher whether this phosphorylation is able to induce transcription or whether transcription factor binding induces phosphorylation by activating the DDR or a non-canonical pathway.

As discussed previously, H2A.X plays a role in the DDR and can be phosphorylated at serine-139 by ATM, ATR or DNA-PK upon checkpoint activation. This could suggest that checkpoint activation, the EMT and chromatin modifications are linked in some manner.

H2A.Z, encoded by *H2AFZ* in humans, has been shown to act as a regulator of the epithelial mesenchymal transition (Domaschenz et al., 2017), discussed in reference to definitive endoderm differentiation. Mice deficient in this protein die early in development, suggesting this is needed during embryonic development (Faast et al., 2001). Many histone variants have been shown to be required in differentiation, H3.3 has been shown to be needed for maintenance of the pluripotent state in ESCs (reviewed by Meshorer and Misteli, 2006). It is also interesting to note that 'poised' gene promoters, thought to be particularly important in embryonic differentiation, are often nucleosome-free or marked with unstable nucleosomes H3.3 and H2A.Z (Jin et al., 2009).

1.3.3.2 The role of p53, p63 and p73 in differentiation

p53, p63 and p73 are members of a family of transcription factors, first discovered for acting as tumour suppressors and controlling cellular stress, known to play key roles in the undifferentiated state, initiation of differentiation and EMT. The p53 protein acts to sense and respond to damage by inducing cell cycle arrest or apoptosis through transcription of multiple transcripts, downstream of the checkpoint kinases (Lane, 1992). While this family of proteins

have a high level of sequence similarity, especially in the DNA binding domain, they are not totally redundant. Genetic knockouts of *Trp53* in mice and *TP53* in human cell lines are viable, suggesting that this protein is not required in differentiation (Donehower, 1996). On the contrary, *Xenopus* embryos depleted of p53 cannot complete gastrulation (Wallingford et al., 1997). These differences may be explained at least in part by the availability of family members to take over as no triple knockout in mice or human has been reported.

In somatic cells, the level of p53 is kept low and this is controlled by the E3-ubiquitin ligase MDM2 (Oliner et al., 1992). *MDM2* null mice do not survive development in the embryo, suggesting the importance of the tight control of p53 levels during development, and these mice can be rescued by genetic knockout of *Trp53* (Jones et al., 1995; de Oca Luna et al., 1995). This suggests that p53 must be downregulated during development. The canonical DDR causes phosphorylation of p53 on serine-15, in human cells, usually by ATM kinase (Banin et al., 1998; Canman et al., 1998; Khanna. et al., 1998). However, there are a vast number of PTMs associated with p53 controlling its activity (reviewed by Gu and Zhu, 2012).

The total level of p53 protein in the embryo, and in mES cells, has been shown to be high, whereas that of p63 and p73 vary depending on the culture conditions (Lin et al., 2005; Lutzker and Levine, 1996; Shigeta et al., 2013). Recent work has shown that the p53 family is important for coordination of signalling pathways controlling mesendoderm differentiation by performing triple knockouts in both mouse and human ES cells (Wang et al., 2017). They show that the decision for a cell to differentiate down the mesendoderm pathway is controlled by the p53 family: the p53 family activates specification genes by inducing expression of genes including *Wnt3* and *Fzd1*. WNT goes on to activate TCF3 which, with SMAD 2/3, binds to enhancers of mesendoderm genes and activates their transcription, thus controlling the exit from pluripotency. This clearly shows that the p53 family are required, if redundantly, in embryonic development. Interestingly, this group show that many of the genes upregulated by p53 in mESC differentiation are also transcripts associated with the DDR, including p21 (Wang et al., 2017).

This connection between Wnt signalling and p53 has been previously shown in mESCs, and has been related to the role of p53 in responding to DNA damage (Lee et al., 2010). Phosphorylation of p53 at serine-315 has also been implicated in the control of *Nanog* expression in mESCs, in order to induce differentiation pathways after DNA damage (Lin et al., 2005).

p53 has been shown to play a role in the control of the EMT: the data suggested that p53 binds to the promoter of miR-200c and miR-183 activating their transcription and controlling expression of EMT specific genes (Chang et al., 2011). They also show that overexpression of *TP53* reverts the mesenchymal phenotype to epithelial. Therefore, p53 is needed to initiate the EMT program but evidence suggests that it is then downregulated by MDM2 to allow completion of this transition (Araki et al., 2010). Another study suggested a more direct interaction in the EMT: TWIST1 binds the p53 C-terminus and this allows MDM2 degradation of p53 (Piccinin et al., 2012). Therefore, p53 is upregulated in the ground state but is downregulated during the EMT such that it should decrease throughout endoderm differentiation after its initiation. It is important to note that in these instances it is the total level of p53 protein, and not the post-translational modifications that change.

Interestingly, the *TP53* transcript has been shown to contain an RNA G-quadruplex (Marcel et al., 2011; Wanrooij et al., 2010). The authors suggest that the G4 sequence in intron three of *TP53,* controls splicing choice and therefore the particular p53 isoform produced. Similarly, *NRAS* and *BCL-X* contain RNA G4s (Burley et al., 2017; Kumari et al., 2007).

Finally, p53 has also been shown to be vital in controlling the methylation state of mES cells in the naïve ground state (Tovy et al., 2017). They show that p53 deficient mES cells do not maintain the hypomethylation required of the ground state due to the control of DNMT3A, DNMT3B, TET1 and TET2 enzymes by p53. The lack of p53 causes an imbalance of the 5mC and 5hmC marks, increasing the methylation in mES cells and increasing heterogeneity DNA methylation (Tovy et al., 2017).

1.3.3.3 The role of caspases in differentiation

Recent work has suggested that cellular differentiation and cell death share many features, suggesting a common pathway in cardiac differentiation (Ghiasi et al., 2018). Differentiation to a number of cell types including skeletal muscle cells, erythrocytes and *Drosophila* sperm has been shown to induce caspase activity (Arama et al., 2003; Fernando et al., 2002; Zermati et al., 2001). This study by Ghiasi et al shows an increase in reactive oxygen species (ROS), p53 and APAF-1 cell death protein during differentiation. They also describe membrane potential loss, PARP-1 cleavage and cytochrome *c* release, all indicative of a cell death response (Ghiasi et al., 2018). Each of these features occur in apoptosis but during cardiac differentiation they occur to lower, reversible levels. They suggest that the timing and intensity of these processes are likely to determine whether cell death or differentiation is the outcome.

Other studies have suggested that although ROS increase during differentiation, the ability for cells to deal with this damage is decreased, therefore the ability for cells to process damage may decrease during differentiation (reviewed by Hirano and Tamae, 2012).

1.3.4 The importance of DDR proteins during development

Genetic knockout of *CHEK1* (Liu et al., 2000), *ATR* (de Klein et al., 2000), *BRCA1*, *BRCA2* and *RAD51* (Lim and Hasty, 1996) often leads to embryonic lethality in mice. Whereas genetic knockout of *TP53* (Donehower, 1996), *ATM* (Barlow et al., 1996) and *CDKN1A* (Martín-Caballero et al., 2001) lead to spontaneous tumour formation. While this does not immediately link these DDR proteins to being essential during development, it does suggest that they could play a role in unchallenged differentiation. Interestingly, the *Rad51* knockout mouse develops further upon additional knockout of *Trp53* suggesting that these proteins need to be kept in check during differentiation (Lim and Hasty, 1996).

Finally, the DSBs have been shown to control the expression of neuronal early-response genes (Madabhushi et al., 2015). This activity-dependent DSB formation is suggested to be induced by Topoisomerase IIβ relieving topological constraints. This highlights the importance of the DDR on gene expression and the interplay of these fields.

1.3.5 Replication stress in development

1.3.5.1 Ultrafine anaphase bridges

Regions in the genome which are hard to replicate can cause the generation of ultrafine anaphase bridges (UFBs). These UFBs need to be resolved for correct chromosome segregation at cell division and this occurs using the protein PICH (Biebricher et al., 2013). They normally form at locations of repetition in the DNA, such as rRNA, telomeres and centromeres, or fragile sites (Nielsen and Hickson, 2016). Unpublished work by Kamikawa and Tsubouchi looking at UFBs in mES cells showed that there were actually fewer in these ES cells, suggesting again that these cells are able to deal with fast replication accurately. Mouse knockouts of *Pich* show that it is embryonic lethal, meaning that the removal of these structures occurs frequently in ES cells and it is necessary for them to be removed (Eliene Albers et al., 2018).

1.3.5.2 DNA origin firing

Due to the rapid cycling of embryonic stem cells they must have tightly regulated control of DNA replication. Origin licencing is under tight control in ES cells to help maintain pluripotency. In human ES cells the G1 phase of the cell cycle is very quick compared to

somatic cells, around 2.5 hours, and this has been suggested to maintain pluripotency (Kareta et al., 2015; Soufi and Dalton, 2016). G1 phase is also when cells respond to differentiation cues, therefore ES cells must integrate numerous signals quickly at this point (Gonzales et al., 2015). A study on the licensing factor MCM loading onto DNA has shown that rapid loading is needed for pluripotency and prevents differentiation (Matson et al., 2017).

MCM2-7 complexes are loaded onto chromatin in excess and these excess complexes normally do not get fired and are therefore called dormant origins (Ge et al., 2007). However, during replication stress these provide a backup for the cell. These dormant origins are considered to play a role in defence against this stress in ES cells as many more of these are recruited (Ge et al., 2015).

1.3.5.3 Replication fork speed changes during differentiation

Throughout cellular differentiation many aspects of DNA replication are altered including changes to origin choice and timing. However, it has also been found that the global fork speed increases dramatically on erythroid differentiation (Hwang et al., 2016). The authors show that this switch is mediated by CDK activity and this causes an increase in fork speed and a decrease in the length of S phase.

1.4 Aims

The aim of this thesis was to be able to understand whether replication impediments could impact differentiation from embryonic stem cells. Previous work in the group has shown that replication impediments can perturb the deposition of histone marks onto the appropriate regions of the DNA and this causes changes to gene expression. However, since all of this work so far has been performed in the terminally differentiated chicken DT40 cell line it is important to show whether this model is consistent with other systems, especially cellular differentiation.

Embryonic differentiation is an incredibly important process and it is kept in check by a large number of mechanisms in very varied pathways. Therefore, it could be that the cell has multiple mechanisms for dealing with replication stress and that this same phenomenon is not seen during the differentiation of ES cells. However, it is compelling to suggest that there may be a role due to the many changes occurring, especially epigenetically in the control of gene expression, during this process.

Initially I set up systems to differentiate both mES and human iPS cells. These differentiation protocols were chosen due to a combination of high efficiency, knowledge of the pathways controlling differentiation, defined protocols and a sufficient number of cell cycles to reveal this replication dependent phenomenon. After validation of the systems I focussed on the hiPSC model of differentiation to definitive endoderm which yielded a high differentiation efficiency. I mainly monitored the differentiation efficiency using permeabilised flow cytometry which gave a clear readout of differentiation on a single-cell basis.

In order to perturb replication, I used ligands that stabilise a specific secondary structure in the DNA, the G-quadruplex, DNA damage inducing agents, nucleotide pool depletion and iPSCs with genetic knockouts of enzymes known to help in the processing of G-quadruplexes in the DNA.

Throughout this thesis I have shown that each of these individual factors was able to alter the differentiation to definitive endoderm. Each altered the differentiation efficiency, level of cell death, and activation of DDR factors to varying extents. I observed very different responses in the undifferentiated state compared to during differentiation when replication was altered. These changes were probed in further detail using RNA sequencing to enable a greater understanding of the pathways affected.

DNA damage in the form of UV irradiation or MMS treatment decreased the efficiency of differentiation while showing a block in G2/M phase of the cell cycle: the efficiency of differentiation negatively correlated with the dose of DNA damaging agent. Interestingly, the analysis of inflicting damage throughout differentiation led me to understand that changes to the levels of proteins in the DNA damage response occur during differentiation. When DNA damaging agents were added, the levels of p53 increased resulting in increased cell death and a lower efficiency of differentiation.

I have also shown that G-quadruplex-binding ligands were able to alter the course of differentiation, either by a p53-dependent mechanism or by a separate, possibly G-quadruplex specific, mechanism. This also occurred in *REV1* knockout cells, a polymerase known to play a role in G-quadruplex processing. G-quadruplex stabilisation during differentiation prevented definitive endoderm specification. I also suggest that the G4 ligand Pyridostatin inhibits differentiation in a manner specific to inducing DNA damage, and that this is not the case with an alternative ligand, PhenDC3.

Chapter 2. Materials and Methods

2.1 Cell Culture

2.1.1 Mouse embryonic stem cell culture

2.1.1.1 Basic cell culture

The cells used were a kind gift from Edith Heard who had analysed X-chromosome inactivation during embryogenesis by differentiating these cells from the ground state to the epiblast stem cell stage (EpiSC) (Schulz et al., 2014). PGK12.1 cells were cultured in DMEM medium supplemented with 15% decomplemented ES -FBS (Gibco), 0.1 mM β-2-mercaptoethanol (Sigma), 1000 U/mL mouse leukaemia inhibitory factor (LIF) (Miltenyi) and PenStrep. E14 cells were maintained in Glasgow medium supplemented with 15% decomplemented ES-FBS (Gibco), 0.1 mM β-2-mercaptoethanol (Sigma), 1000 U/ mL mouse leukaemia inhibitory factor (LIF) (Miltenyi), 0.1 mM MEM non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 2mM Glutamax and PenStrep. Cells were grown in an incubator (37°C, 5% CO_2) on 24 well or six-well plates (Costar) coated in 0.1% gelatin (Sigma) in 2 – 4 mL medium and passaged every two-three days. Cells were washed with 1 mL PBS before splitting and then incubated at 37°C (5 min) with 1 mL ES-trypsin. Cells were resuspended in medium and plated. Cell medium was replaced every day.

2.1.1.2 Two inhibitor culture

Cells were adapted to LIF and 2i culture medium over a number of passages by changing the medium and splitting using accutase (Gibco) instead of trypsin. Medium consisted of N2B27 supplemented with 1000 U/mL LIF, 3 µM GSK3 inhibitor CHIR99021 (Tocris) and 1 µM MEK inhibitor PD0325901 (Syn Kinase). N2B27 medium contained 50% Neurobasal medium (Gibco), 50% DMEM/F12 (Gibco) supplemented with 2 mM Glutamax (Gibco), 0.1 mM β-2-mercaptoethanol, NDiff Neuro2 supplement (Millipore) and B27 serum free supplement (Gibco).

2.1.1.3 Differentiation

In order to induce differentiation, mES cells cultured on LIF and 2i were passaged onto plates coated in Fibronectin (10 µg/mL) in N2B27 media supplemented with 10 ng/mL FGF2 (R&D) and 20 ng/mL Activin A (R&D). Initial cell density was around 8×10^4 cells/cm^2.

2.1.1.4 Freezing and thawing

Frozen cells were thawed in the bead bath until just a small ice crystal was left and then added to 10 mL warm serum containing LIF medium. Cells were centrifuged (290 *g*, 5 min), supernatant was removed, and cells were resuspended in 5 mL medium. Cells were plated as serial dilutions onto precoated gelatin-coated 24 well plates and put into the incubator. Cells were frozen from 24 well plates by removing cell culture medium, washing with PBS and adding 350 µL ES-trypsin. Cells were put into the incubator until all clumps were single cells (2-5 min) and checked under the microscope. Cells were resuspended and added to 150 µL medium without LIF and transferred to a cryovial, 500 µL 2X freezing medium (90% ES-FBS/10% DMSO) was added and the cells were frozen slowly at -80ºC using Mr. Frosty™ freezing container. After eight hours the cells were transferred to the liquid nitrogen for long term storage.

2.1.1.5 Mycoplasma testing preparation

Mycoplasma testing was performed once a month by the LMB mycoplasma testing service in all mouse and human cell lines. Mouse lines were grown without antibiotic for three days before testing. One mL of spent cell culture media that had not been changed for at least 24 hours was removed from three wells of each cell line, placed in a 15 mL tube and centrifuged (200 *g*, 5 min). One mL of supernatant per sample was placed in a sterile Eppendorf® and stored at 4°C until testing. Cells remained mycoplasma negative throughout the project.

2.1.2 Human induced pluripotent stem cell culture

2.1.2.1 Basic cell culture

The BOBSC cells (Yusa et al., 2011) were obtained from the Sanger Centre. Knockout BOBSC lines were generated as part of the COMSIG (Causes of Mutational SIGnatures) project, funded by the Wellcome Trust, using a CRISPR/Cas9 strategy in which one allele was disrupted by gene targeting and the other by error-prone repair (Table 1). All cells, including knockouts, were cultured in Essential 8™ or Essential 8™ Flex medium on six-well plates coated in Vitronectin-XF (Stem Cell) kept at 37ºC, 5% CO_2 unless otherwise stated. Cells were passaged 1:10 every three or four days depending on confluency. Cell media was aspirated and one mL of 0.5 mM EDTA (Thermo Fisher Scientific) was washed over the cells. The cells were left at room temperature in fresh EDTA for five minutes before the EDTA was removed and the cells were blasted with media to remove them from the plate before being transferred to a new plate. Cells were maintained as small clumps.

Cell line	Code	Barcode	Targeting
Wildtype BOBSC	T6/8 B1	BOBSC-T6/8_B1	n/a
REV1 KO	HUFP0046_1_A_B06	1095802743	Exon 12
WRN KO	HUFP0034_1_A_E06	1095778295	Exon 12
PRIMPOL KO	HUFP0046_1_A_C02	1095802750	Exon 5
TP53 KO	HUFP0007_3_A_H04	1095765195	Exon 6
ATM KO	HUFP0029_1_B_A12	1095717002	Exon 7

Table 1. hiPS cell lines used throughout the project, the knockout cell lines were generated by the Sanger Centre COMSIG project. Specific targeting information can be found in: https://www.sanger.ac.uk/htgt/lims2/public_reports/well_genotyping_info_search.

2.1.2.2 Endoderm differentiation

Cells were passaged 1:8 onto vitronectin coated six-well plates one day prior to setting up endoderm differentiation. On day one of differentiation, cell media was changed to CDM-PVA supplemented with 100 ng/mL Activin A (R&D), 80 ng/mL FGF2 (R&D), 10 ng/mL BMP4 (R&D), 10 µM PI3K inhibitor LY294003 (Promega) and 3 µM GSK3i CHIR99021 (Tocris) (initially Selleckchem, but this was found to be unreliable). On day two, cell medium was replaced with CDM-PVA supplemented with Activin A, FGF2, BMP4 and LY294003 as above. CDM-PVA consisted of 50% Ham's F-12 (Gibco) and 50% IMDM (Gibco) supplemented with 1 g/L PVA (Sigma), 1 mM concentrated Lipids (Life technologies), 0.5 mM Thioglycerol (Sigma), 15 µg/mL Transferrin (Roche) and 7 µg/mL Insulin (Roche). The PVA solution was first made up in MPW by heating to 90ºC whilst stirring. The stirring was continued overnight, while cooling, to ensure the powder was dissolved. On day three, the media was replaced with RPMI+ supplemented with 100 ng/mL Activin A and 80 ng/mL FGF2. RPMI was supplemented with 1 mM NEM-NEEA (Gibco) and B27 supplement. Recombinant proteins were added at concentrations stated in the main text: Cerberus 1 (R&D), Lefty-A (R&D).

2.1.2.3 Neuroectoderm differentiation

Cells were set up for neuroectoderm differentiation in a similar manner to definitive endoderm differentiation, cell media was changed every day. On day one and two cells were cultured in CDM-PVA, as above, supplemented in FGF2 (20 ng/mL), CHIR99021 (3 µM), LDN-193189 (0.1 µM) (EZSolution Source Bioscience) (Selleckchem was used initially but yielded poor results) and SB431542 (10 µM) (Tocris). From day three to six, cells were cultured in N2B27 medium, as before, with 10 µM SB431542. From day seven onwards, cells were kept in N2B27 until analysis.

2.1.2.4 Freezing and thawing

hiPSCs were thawed as above for mESCs but resuspended in 3 mL Essential 8™ medium and plated on precoated vitronectin six-well plates at 1 mL per plate. One well of cells was frozen by removing medium, washing with EDTA and adding 1 mL EDTA (5 min, RT). EDTA was aspirated, clumps of cells were resuspended in 500 µL DMEM medium and 500 µL 2X freezing medium was added as in mES cells. Cells were frozen in Mr. Frosty™ and then transferred to liquid nitrogen.

2.1.3 Human embryonic stem cell culture

The H9 hES cells were provided by the lab of Madeline Lancaster (MRC LMB). They were cultured, frozen, thawed and differentiated as above for BOBSC cells.

2.1.4 Treatment of cells

2.1.4.1 G-quadruplex-binding ligands

When cells were cultured in the presence of G4 ligands, the ligands were thawed, vortexed due to their poor solubility and added to medium. This ligand-containing medium was vortexed again before being added to cells. If the cells were to be kept in ligands for over a week the medium was filtered to prevent contamination.

2.1.4.2 DNA damaging agents

DNA damaging agents were added as for ligands, but in the case of MMS it was first diluted 1000X in cell medium and vortexed before being added to cells at the correct concentration.

2.1.4.3 UV irradiation

The UV-C source was turned on 30 min prior to use to stabilise. Bulb output was assessed with a calibrated UV-C meter (UVP Inc) and length of time for a given dose was calculated (0.5-20 J/cm^2). Cells media was removed and replaced with 1 mL PBS per well. The lid of the plate was removed, and cells were transferred to the UV box, the shutter was opened for the time required at that given dose and then shut. Cell media was replaced, and cells were returned to the incubator. This was only performed in six-well plates.

2.1.4.4 Cell counting

Both mouse and human cells were counted using the Countess Cell Counter (Thermo Fisher) ‘ES’ cell program to analyse the viability and number of cells.

2.1.4.5 Dead cell removal

Dead cells were removed from non-adherent samples using the MACS dead cell removal kit (Miltenyi Biotec). Cells were centrifuged (300 *g*, 5 min), resuspended in 100 µL microbeads and incubated (RT, 15 min). The MACS columns were put into the magnetic rack and 500 µL binding buffer was run through the column. Once most of the buffer had eluted, a new 15 mL falcon tube was placed under the column. 500 µL binding buffer was added to the cells and magnetic beads and this was added to the column. The flow through was collected in the falcon tube and the column was washed three times with 500 µL binding buffer to collect the last cells. Binding buffer (500 µL) was added to the column and flushed with the plunger, and the eluate was thrown away. The tube containing the alive cells was spun down (300 *g*, 5 min) and collected for permeabilised flow cytometry, Section 2.3.

2.2 Western blotting

2.2.1 Protein extraction

Medium from one well of a six-well plate was aspirated and one millilitre of single cell dissociation buffer (Gibco) was added (37ºC, 10 min). Cells were transferred to an Eppendorf®, centrifuged (1500 *g*, 4 min) and aspirated, leaving the cell pellet. Cells were washed twice in 500 µL PBS and then resuspended in 100 µL roughly per 1×10^6 cells RIPA buffer (Cell Signalling) containing cOmplete™ protease inhibitor (Roche) and phosphatase inhibitor Halt™ (Thermo Scientific) 1X. Tubes containing cells and buffer were rotated slowly (4ºC, 1 h) and then transferred to a cold centrifuge (16,000 *g*, 4ºC). The supernatant (protein) was transferred to a clean tube, care was taken not to disturb the DNA pellet at the bottom. 5X Laemmli buffer (Sigma) was added and the protein was boiled (95ºC, 5 min). Protein was stored at -20ºC until use.

2.2.2 SDS-PAGE gel electrophoresis

2.2.2.1 Bis-Tris gel

Protein was thawed on ice until liquid and was then run on a NuPAGE® Bis-Tris 4-12% precast gel (Thermo Fisher) unless stated otherwise. The gel was set up and 1X MOPS buffer (50 mM MOPS, 50 mM Tris, 0.1% SDS, 1 mM EDTA, pH 7.7) was added (300 mL) to fill the central well and half of the remaining tank. 10 µL PageRuler™ prestained protein ladder (Thermo Fisher) was added to one well per gel and 20 µL of each protein sample was added to the remaining wells. The gel was electrophoresised (120 V, 2 h) until the 10 kDa marker had almost

reached the bottom of the gel. The gel was removed from the casing and put into DI water until the transfer.

2.2.2.2 Tris-Acetate gel

Tris-Acetate gels were run when the proteins were to be transferred using a wet transfer system. The system was set up as above but using a NuPAGE® 3-8% tris acetate gel and run in NuPAGE® 1% tris acetate SDS running buffer. The gel was either run as above or for ATM (40 V, 4ºC, overnight).

2.2.3 Protein transfer

2.2.3.1 Semi-dry transfer

Semi-dry transfers were performed using the iBlot® 2 system (Thermo Fisher) using iBlot® 2 transfer stacks. Gels were put into the transfer cassette on top of the nitrocellulose membrane followed by two pieces of wet Whatman™ paper, bubbles were removed, and the top stack added. The iBlot® was sealed and proteins were transferred (25 V, 7 min) to the nitrocellulose membrane. The membrane was put into deionised water, cut to size and stained with Ponceau S (15 s, RT, rotating). The membrane was washed twice and blocked.

2.2.3.2 Wet transfer

Wet transfer was only performed for proteins that did not give a signal using the semi-dry system, including phospho-p53 and ATM. The XCell II™ Blot Module was used for the transfer process and performed in NuPAGE® Transfer Buffer containing 10% methanol. The system was set up as in in the manual, with a nitrocellulose membrane for transfer and proteins were transferred (25 V, 2 h, 4ºC).

2.2.4 Western blotting

2.2.4.1 Blocking

The membrane was blocked in 5% milk (Marvel) in 1X TBST (150 mM NaCl, Tris HCl pH7.4 10 mM, 0.1% tween) (1 h, RT, rotating). The membrane was then probed with antibodies.

2.2.4.2 Antibody staining

Primary antibodies were added as stated in Table 2 below. Membranes were incubated with primary antibody overnight, rotating at 4°C unless otherwise stated. The membrane was washed three times for 5-15 minutes while rotating at room temperature with 1X TBST. The membrane

was then blotted with secondary antibody (1 h, RT, rotating). All secondary antibodies were HRP-conjugated (Dako P0447-9) used at 1:5000 dilution. The secondary antibody was washed as for the primary antibody and then the HRP was visualised.

2.2.4.3 Visualisation

The membrane was transferred to a clean piece of plastic and 3 mL HRP substrate (Millipore Luminata Crescendo) was added for five minutes then removed and the membrane was transferred to a cassette. The blot was revealed in the dark room for five seconds to 15 minutes, depending on the strength of the protein band, and then processed on a film developer. Standard protein sizes were marked on and bands were compared to expected sizes.

Protein epitope	**Dilution**	**Catalogue no.**	**Manufacturer**	**Species**
γH2A.X (ser139) JBW301	1:5000	05-636	Merck Millipore	Mouse
p53 (DO-1)	1:1000	Ab1101	Abcam	Mouse
p-p53 Ser15	1:1000	9284	CST	Mouse
PC10	1:5000	SC-56	Santa Cruz	Mouse
TUBULIN (B512)	1:10000	T6074	Sigma	Mouse
β-ACTIN	1:20000	Ab8227	Abcam	Mouse
pCHK1 Ser345 (133D3)	1:1000	113D3	CST	Rabbit
CHK1	1:1000	Ab40866	Abcam	Mouse
pRPA32 (S33)	1:1000	A300-246A	Bethyl	Rabbit
RPA32	1:2000	A300-244A	Bethyl	Rabbit
H2A.X total	1:1000	2595S	CST	Rabbit
MDM2 (SMP14)	1:200	SC-965	Santa Cruz	Mouse
pCHK2	1:1000	2661	CST	Rabbit
CHK2	1:50000	Ab109413	Abcam	Rabbit
ATM	1:5000	Ab17995	Abcam	Rabbit

Table 2. A list of all primary antibodies used for western blotting.

2.3 Flow cytometry

All cells were sorted for collection on the Beckman Coulter MoFlo™ by Maria Daly and the Flow Cytometry team at the LMB. Unless stated otherwise, flow cytometry was performed on the BD LSRFortessa™, which is equipped with 405 nm, 488 nm, 561 nm and 640 nm lasers. Cells were monitored for GFP or antibody staining, combined with DAPI and analysed using FlowJo®, LLC. Where possible 100,000 cells were analysed.

2.3.1 GFP monitoring

In the GFP-tagged mESCs, cells were analysed for GFP in the LIF, 2i and EpiLC states. Cells were washed with PBS and incubated with ES-trypsin at 37ºC until the cells were in a single cell suspension. The cells were transferred to a 15 mL falcon tube with 5 mL warm cell culture medium, centrifuged (300 *g*, 5 min), washed with 5 mL PBS and spun down again. The cell pellet was resuspended in 500 µL warm medium and transferred to flow cytometry tubes (Sarstedt) to be analysed for GFP.

2.3.2 Permeabilised flow cytometry

Around 1×10^6 cells were collected to be analysed by permeabilised flow cytometry for each condition. Cell media was aspirated and one millilitre of Cell Dissociation Buffer (Thermo Fisher Scientific) was added for ten minutes at 37ºC. The cells were collected, transferred to an Eppendorf® and centrifuged (1500 *g*, 4 min). The cell pellet was fixed in 200 µL 1% PFA for human cells or 4% PFA for mESCs (10 min, RT) and centrifuged (1500 *g*, 4 min). The pellet was resuspended in 200 µL of 90% PBS/10% DMSO and kept at -80ºC until preparation for flow cytometry.

Fixed cells were thawed at room temperature for ten minutes, spun down (1500 *g*, 4 min), aspirated and resuspended in 200 µL 1XBD Perm/Wash™ buffer. Half of the volume of a positive control sample was transferred into a new tube for the IgG control, all tubes centrifuged (1500 *g*, 4 min), and resuspended in 100 µL BD buffer. Cells were permeabilised and blocked (15 min, RT) and spun down for antibody staining: either primary followed by secondary or conjugated antibody staining (see below). Mouse cells were permeabilised in 0.1% PBS-tween (20 min, RT) and then blocked in 0.1% BSA/10% normal goat serum/0.3 M glycine (1 h, RT) prior to antibody staining (see below).

2.3.2.1 Primary and secondary antibody staining

Cells were resuspended in 100 μL anti-γ-H2A.X antibody (1:500) (Merck Millipore 05-636) (ab2893 and ab11174 Abcam rabbit antibodies were also used when combining staining with other mouse antibody combinations) (4°C, overnight) or IgG control without antibody. Cells were spun down (1500 *g*, 4 min), washed in BD buffer twice and resuspended in secondary antibody (1:200) (1 h, RT, dark). Cells were spun down and washed, as above, and resuspended in 400 μL PBS/BSA 0.5% with 1 μg DAPI per sample. Cells were transferred to flow cytometry tubes and covered from light until analysis on the BD LSRFortessa™.

A number of other antibodies were used in a similar manner as shown in Table 3 below; all primary antibodies were used at 1:500 and all secondaries (Invitrogen Alexa Fluor®) at 1:200 unless otherwise stated.

Protein	**Primary Antibody**	**Secondary Antibody**
γH2A.X (ser139)	Merck Millipore 05-636 Ms Abcam ab2893 Rb Abcam ab11174 Rb	Alexa Fluor® 488 A10680 Gt α Ms Alexa Fluor® 594 A21207 Dk α Rb Alexa Fluor® 594 A21207 Dk α Rb
H2A.X	Cell Signaling (CST) 2595 Rb	Alexa Fluor® 594 A21207 Dk α Rb
p53	Santa Cruz DO-1 sc-126 Ms	Alexa Fluor® 488 A10680 Gt α Ms
REX1	Abcam ab28141 Rb	Alexa Fluor® 568 A11011 Gt α Rb
DNMT3A	Abcam ab13888 Ms	Alexa Fluor® 647 A21234 Gt α Ms
pRPA32 (S33)	Bethyl A300-246A Rb	Alexa Fluor® 594 A21207 Dk α Rb

Table 3. A list of primary and secondary antibodies used in combination for permeabilised flow cytometry.

2.3.2.2 Conjugated antibody staining

Cells were resuspended in 100 μL BD buffer containing 4 μL conjugated antibody, or appropriate isotype control, per reaction and incubated in the dark (1 h, RT). Cells were washed once in BD buffer, spun down (1500 *g*, 4 min) and resuspended in 400 μL PBS/BSA 0.5% with 1 μg DAPI per sample. Cells were transferred to flow cytometry tubes and analysed as above.

Protein	Conjugated Primary Antibody
SOX17	BD Pharmingen™ 562205 Alexa Fluor® 488 Ms
EOMES	Invitrogen 50-4877-42 eBioscience™ eFluor® 660 Ms
p53	Santa Cruz P53 DO-1 sc-126-PE PE conjugated
PAX-6	BD Pharmingen™Alexa Fluor® 647 Ms 562249
SOX1	BD Pharmingen™ 561549 PerCP-Cy™5.5 Ms

Table 4. A list of conjugated antibodies used for permeabilised flow cytometry.

2.3.3 Cell death staining

In order to determine whether cells, that had lost adherence to culture plates, were apoptotic, necrotic or still alive, cells were stained using the FITC Annexin V Apoptosis Detection Kit with 7-AAD (Biolegend®). Cell supernatant or adherent cells were collected, transferred to a falcon tube and spun down (300 *g*, 4 min). Cells were washed twice in Cell Staining Buffer and then resuspended in 100 µL Annexin V Binding Buffer per well of a six-well plate. The cells were added to a flow cytometry tube, 5 µL FITC Annexin V and 5 µL 7-AAD was added. Cells were vortexed and incubated in the dark (RT, 15 min). 400 µL Annexin V Binding Buffer was added to each tube and the cells were analysed using flow cytometry.

2.3.4 Live cell sorting

Mouse ES cells were sorted for live cells expressing GFP by Maria Daly and the MRC LMB flow cytometry team. Cells were washed with PBS, trypsin was added (37ºC, 2-5 min) until single cells could be seen. Cells were added to 5 mL warm culture medium, spun down (290 *g*, 7 min), washed in PBS and spun down again. Cells were resuspended in 1 mL warm medium and 1 mL PI solution (15 µg/mL PI in 2.5% FBS/PBS). The solution was transferred through a strainer to remove clumps into a sorting tube and taken to be sorted for GFP positive, PI negative populations.

2.4 Cell cycle Analysis

2.4.1 EdU visualisation

2.4.1.1 EdU Click-iT® kit

The kit was prepared as stated in the EdU Click-iT® protocol but the EdU was made up in MPW instead of DMSO.

2.4.1.2 EdU staining

Cells at around 30% confluency had 25 µM EdU per well of a six-well plate (or adjusted accordingly) added to cell culture medium and incubated for one hour at 37ºC. The cells were then collected and fixed with 1% PFA as above in the permeabilised flow cytometry Section 2.3.2.

2.4.1.3 Click-iT® reaction

Cells were thawed at room temperature until liquid, spun down (1500 *g*, 4 min) and resuspended in 100 µL BD buffer (above). Cells were centrifuged again and resuspended in 100 µL BD buffer (RT, 15 min) to permeabilise and block. The cells were spun down and resuspended in 125 µL Click-iT® reaction buffer as in the kit (containing $CuSO_4$ and fluorescent azide) for 30 minutes in the dark at room temperature. The cells were then centrifuged, and the pellet resuspended in 1.5 µg RNaseA, 1 µg DAPI in BD buffer per condition and left (RT, 30 min or overnight, 4ºC) until the cells were analysed by flow cytometry. Cells could also be stained for permeabilised flow cytometry before being resuspended in the final buffer.

2.4.2 Cell division mapping

To trace the number of cell cycles that had occurred, the CellTrace™ Violet Cell Proliferation Kit (Invitrogen™) was used.

2.4.2.1 Set up of CellTrace™

20 µL of DMSO was added to the tube of fluorophore and this was added to 20 mL of warm PBS to make a 5 µM solution. Cell media was removed from each well of a six-well plate and cells were incubated with 1.5 mL of solution per well (37ºC, 20 min). The cells were washed twice with warm media and then returned to the incubator until analysis. Control cells were collected just after setting up to compare the initial fluorescence.

2.4.2.2 Visualisation

Cells were collected as with permeabilised flow cytometry and stored until analysis in order to analyse multiple timepoints and conditions at once. Cells could be permeabilised to combine with any permeabilised flow cytometry staining and then analysed using flow cytometry.

2.4.2.3 Analysis

Cells should lose half of the fluorescence for each cell division as this is proportional to the amount of dye in each cell. Therefore, comparing the fluorescence of the cells collected at the start and end of differentiation allowed the number of cell divisions to be quantified using Equation 1.

$$N = \frac{\log\left(\frac{Co}{Ct}\right)}{\log 2}$$

Equation 1. The formula for the number of cell divisions (N) in a CellTrace™ experiment where C_0 is the fluorescence at time 0 and C_t is the fluorescence after N cell divisions.

2.4.3 Quantification using FlowJo® software

Cells collected for permeabilised flow cytometry were stained with DAPI to analyse the DNA content of the cells using the violet V-450 laser. The population of cells in each phase of the cell cycle was quantified using the 'Cell Cycle' function in the FlowJo® flow cytometry analysis software. This gives a crude representation of the percentage of cells in G1, S or G2/M phase.

2.4.4 Cellular synchronisation

Cells were blocked in prometaphase using 200 ng/mL nocodazole (Sigma) for 16 hours. Cells were released by washing once with DPBS and twice with warm medium. Cells were then either differentiated or kept in the undifferentiated state.

2.5 RNA analysis

2.5.1 RNA extraction

Before RNA extraction, pipettes and bench space were cleaned with RNaseZAP® (Thermo Fischer) to remove any RNase contamination and filter tips were used throughout the extraction process. RNA extraction was performed using the RNeasy® Qiagen RNA extraction kit. Medium was aspirated from adherent cells and replaced with 600 µL RLT buffer containing β-mercaptoethanol. Cells were resuspended in buffer and transferred to sterile Eppendorf® tubes. At this stage cells could be kept at -80ºC or extracted for RNA immediately. Extraction was performed as stated in the protocol but without the initial spin to remove debris. RNA was eluted into 30 µL nuclease free water and 1 µL was analysed on the nanodrop to check the RNA concentration and for contamination with DNA or protein. RNA was stored at -80ºC.

2.5.2 Reverse transcription

RNA was reverse transcribed to DNA using Qiagen QuantiTect® reverse transcription kit using 800 ng RNA. The initial gDNA wipeout was performed (42ºC, 2 min) and the reverse transcription (42ºC, 25 min). cDNA was stored at -20ºC for up to 48 hours before quantitative (qPCR) was performed.

2.5.3 Quantitative PCR

2.5.3.1 Primer design

Primers for qPCR were designed using the Primer3web software with an amplicon size of 100-180bp and an optimum melting temperature of 60°C for control, pluripotency, mESC and EpiLC transcripts as well as hiPS cell differentiation transcripts. Where possible, primers were designed to amplify across intronic regions to prevent gDNA contamination, biasing the results. Primer sequences are shown in Table 5, Table 6 and Table 7.

2.5.3.2 Primer validation

Primers were validated using qPCR amplification across a range of cDNA dilutions: 1:4, 1:16, 1:64, 1:256 and a no template control. The qPCR mastermix was made up of 50% SYBR® Green Mastermix (Applied Biosystems), 10 µM forward (Fwd) and 10 µM reverse (Rv) primer and 10% diluted cDNA in DEPC. Reactions were set up on a 96 or 384-well optical plate in 20 µL volumes in triplicate and sealed with optical film. The plate was spun down (3220 *g*, 1 min) and run on a ViiA7™ real-time system for 45 cycles (hold stage: 50ºC, 2 min and 95ºC, 10 min and PCR stage: 95ºC, 15 s and 60ºC, 1 min). The melt curve was analysed for each primer pair to check the amplification and the efficiency of the primers was calculated using Equation 2 below; if the efficiency was between 90-105 then the primers were able to be used in experiments.

$$\boldsymbol{Efficiency} = 1 + 2^{\frac{-1}{dy/dx}} \times 100 \qquad \textbf{90-105\%}$$

$$\boldsymbol{R^2 > 0.98}$$

Equation 2. Used to determine the efficiency of a set of qPCR primers. If the efficiency was between 90 and 105% and the R^2 value above 0.98 then the primers could be used to quantify gene expression.

2.5.3.3 Primers for qPCR

The primers used in mouse and human cell lines are shown in Table 5, Table 6 and Table 7.

Transcript	**Fwd Primer: 5'-3'**	**Rv Primer: 5'-3'**
β-actin	CTT TGC AGC TCC TTC GTT GC	CGA TGG AGG GGA ATA CAG CC
Gapdh	GCA TCT TCT TGT GCA GTG CC	ATG AAG GGG TCG TTG ATG TGG C
Oct4/Pou5f1	AGC ACG AGT GGA AAG CAA CT	TCT GCA GGG CTT TCA TGT CC
Nanog	TTC TTG CTT ACA AGG GTC TGC	AGA GGA AGG GCG AGG AGA
Sox2	GCG GAG TGG AAA CTT TTG TCC	GGG AAG CGT GTA CTT ATC CTT CT
Gbx2	GCT TTC TCT GCG GCC GAA	AAG CTC TCC TCC TTG CCC TT
Rex1/Zfp42	TGG GTA CGA GTG GCA GTT TC	CCA CGT GTC CCA GCT CTT AG
Stella/Dppa3	GAC CCA ATG AAG GAC CCT GAA	GCT TGA CAC CGG GGT TTA G
T/Brachyury	CTC TCT CTC CCC TCC ACA CA	ACT GCA GCA TGG ACA GAC AA
Nodal	TCA AGC CTG TTG GGC TCT ACT	GTC AAA CGT GAA AGT CCA GTT CT
Fgf5	ACG AGG AGT TTT CAG CAA CA	CCA CTC TCG GCC TGT CTT TT

Table 5. A list of qPCR primers to amplify mouse transcripts during EpiLC differentiation.

Transcript	**Fwd Primer: 5'-3'**	**Rv Primer: 5'-3'**
PBGD	GGA GCCA TGT CTG GTA ACG G	CCA CGC GAA TCA CTC TCA TCT
GAPDH	TCA CCA GGG CTG CTT TTA ACT	GAC GGT GCC ATG GAA TTT GC
β-ACTIN	CGC GAG AAG ATG ACC CAG AT	ATC ACG ATG CCA GTG GTA CG
HPRT1	AGG CTT TGG ACG GCC TCT GGA A	CGA ATG ACA CCG TAC TCC TCA TAG AAG CT
POU5F1	AGT GAG AGG CAA CCT GGA GA	ACA CTC GGA CCA CAT CCT TC
SOX2	TGG ACA GTT ACG CGC ACA T	CGA GTA GGA CAT GCT GTA GGT
NANOG	CAT GAG TGT GGA TCC AGC TTG	CCT GAA TAA GCA GAT CCA TGG
FOXH1	GAT CGC CTT GGT GAT TCAG	TTC CAG CCC TCG TAG TCTT C
MIXL1	GGT ACC CCG ACA TCC ACT TG	TAA TCT CCG GCC TAG CCA AA
NODAL	TGA GCC AAC AAG AGG ATC TG	TGG AAA ATC TCA ATG GCA AG
EOMES	ATC ATT ACG AAA CAG GGC AGG C	CGG GGT TGG TAT TTG TGT AAG G
GSC	GAG GAG AAA GTG GAG GTC TGG TT	CTC TGA TGA GGA CCG CTT CTG
T	TGC TTC CCT GAG ACC CAG TT	GAT CAC TTC TTT CCT TTG CAT CAA G
SOX17	CGC ACG GAA TTT GAA CAG TA	GGA TCA GGG ACC TGT CAC AC

Table 6. A list of qPCR primers used to analyse the BOBSC differentiation to definitive endoderm.

Transcript	Fwd Primer: 5'-3'	Rv Primer: 5'-3'
GAPDH	TCA CCA GGG CTG CTT TTA ACT	GAC GGT GCC ATG GAA TTT GC
β-ACTIN	CGC GAG AAG ATG ACC CAG AT	ATC ACG ATG CCA GTG GTA CG
REV1	CGT GGC TTG GAT AGA CCA	ACG TTT ACC CTT CAT GCC AGT
WRN	CCA GCA CCC AAT GAA GAG CA	GCC ATG ACA GCA ACA TTA TCT CT
PRIMPOL	GAG CAA AGC AGT CCT GAC CT	GTA ACC TCC AAA GCC ACA CG

Table 7. A list of qPCR primers used to verify COMSIG BOBSC knockout cell lines.

2.5.3.4 RNA expression quantification

Once the primers had been validated as efficient, cDNA from cells cultured in each condition was diluted 1:5 and the reaction was set up as above for primer validation. Once the C_T values had been obtained, gene expression was normalised to the geometric mean of the control genes, using the GEOMEAN function on Excel, and a relative expression value produced.

2.6 Next generation sequencing

2.6.1 RNA sequencing

2.6.1.1 RNA extraction

RNA was extracted as in Section 2.3.1 and then quantified using the nanodrop to verify that the concentration was high enough to use and that there was no protein or DNA contamination. RNA was collected from separate samples on the same day to minimise variation in culture conditions, while this makes it easier to analyse this leads to some caveats in understanding the significance of specific numbers of genes, the broad picture should be the same.

2.6.1.2 RNA quality analysis

To check the quality of RNA it was run on an Agilent RNA Pico 6000 chip using an Agilent 2100 Bioanalyzer. The RIN score was deduced and RNA with a score above seven was used to generate RNA libraries.

2.6.1.3 RNA library preparation

750 ng of RNA was diluted into 50 µL DEPC water and kept on ice. RNA libraries were prepared using the NEBNextUltraII RNA library preparation kit (E7770) and PolyA tail isolation with eight PCR cycles and the NEBNext Oligos 1-24 and then stored at -20ºC until analysis.

2.6.1.4 RNA library quality assessment

One µg of each prepared library was run on an Agilent 2100 Bioanalyzer High Sensitivity DNA chip and the electrogram was analysed for shape of the graph and the average size of each library (bp): the library was checked for primer (~80 bp) and adapter (~128 bp) contamination and any libraries with high content were re-purified. Once the libraries were of high enough quality they were quantified.

2.6.1.5 Library quantification

Libraries were quantified using the KAPA Library Quantification Kit Illumina®. Libraries were serial diluted to 1:100,000 and 1:1,000,000 in 10 mM Tris-HCl pH 8.0 and quantified in triplicate as stated in the protocol using the Rox-low buffer for the ViiA7 qPCR system. Melt-curves for the libraries were also analysed to assess the quality of the libraries. Average C_Ts were transferred to the KAPA quantification excel spreadsheet with standard values and the concentration of each library was calculated.

2.6.1.6 Library pooling

After quantification, libraries created with different NEBNext Oligos were pooled to a total concentration of 20 nM in 10 mM Tris-HCl pH 8.0. Between 12 and 18 libraries were pooled for each sequencing lane in order to give at least 100,000 reads per library. The pooled library was run on a High Sensitivity DNA chip, as above, and the average size of the library calculated before sending to sequencing. 20 µL of pooled library was handed to CRUK Genomics Core and sequenced on a HiSeq4000 machine with single end reads.

2.6.2 Chromatin immunoprecipitation (ChIP) sequencing

Protocol for ChIP taken from Davide Schiavone, former member of the Sale lab.

2.6.2.1 Cell collection and crosslinking

hiPS cells were cultured in 15 cm dishes for ChIP and DT40 cells used as a 'spike-in control' were cultured by Benedicte Recolin in the lab and treated with 2 mM HU for 16 hours before collection. When the cells were ready to be collected, one dish of cells was counted and on the other six dishes the medium was removed and replaced with 20 mL DPBS. Plates were taken to the fume hood, 600 µL formaldehyde (final concentration of 1%) (Santa Cruz) was added and plates were incubated (RT, rotating, 10 min). To quench the reaction 2 mL of 2 M glycine was added (0.2 M final concentration) (RT, rotating, 5 min) and the cells were cooled in plates (10 min, 4ºC). Cells were scraped off the plates, transferred to 50 mL falcon tubes and spun

down (400 *g*, 5 min, 4ºC). Formaldehyde-containing medium was disposed of appropriately and the cell pellet was washed with 10 mL cold PBS. Cells were spun down (400 *g*, 5 min, RT) and the washing was repeated. The PBS was removed, the cell pellet was snap frozen in liquid nitrogen and kept at -80ºC until cell lysis was performed.

2.6.2.2 Cell lysis

Cells were thawed on ice, resuspended in 10 mL of ChIP lysis buffer (10 mM tris pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP40, 0.5 mM PMSF), incubated (RT, 5 min) and spun down (200 *g*, 5 min). This was repeated twice, and the cells were either snap frozen as above or continued to be sonicated.

2.6.2.3 Sonication

Cells were resuspended in sonication buffer (1% SDS, 10 mM EDTA, 50 mM tris pH 8, 1X PIC, 0.5 mM PMSF) at 40,000 cells/µL and then kept on ice for 30 minutes with occasional vortexing until sonication. Cells were aliquoted into 1.5 mL tubes with 400-600 µL in each tube, and the same volume in each tube per sonication. Cells were sonicated using the Bioruptor using 30 cycles (30 s on / 30 s off, high power, 4ºC). Cells were removed from the Bioruptor, 100 µL was transferred to a new tube to analyse sonication efficiency and the remainder was snap frozen in liquid nitrogen.

2.6.2.4 Analysis of sonication efficiency

100 µL MPW, 8 µL of 5 M NaCl and 2 µL 10 mg/mL RNase A were added to each of the sheared chromatin samples to analyse the sonication efficiency. The cells were vortexed and incubated with shaking overnight (65ºC, 300 rpm). Tubes were briefly centrifuged, 10 µL of 10 mg/mL Proteinase K was added and the cells were incubated (42ºC, 1.5 h). The samples were then electrophoresed with loading dye on an 1.5% agarose gel (90 V, 1.5 h) with 1 Kb plus DNA ladder (Thermo Fisher). The DNA fragment size was analysed and checked for 500 bp fragments.

2.6.2.5 Immunoprecipitation

Chromatin was thawed, diluted in 3 mL dilution solution (1.1% triton X-100, 1.2 mM EDTA, 16.7 mM tris pH 8, 167 mM NaCl, 1X PIC, 0.5 mM PMSF) and the human chromatin was spiked with 1:10 DT40 chromatin. 50 µL was removed and stored at -20ºC for total input sequencing. The diluted chromatin was aliquoted at 1 mL per 1.5 mL tube and 2 µg of antibody was added: H3 (Abcam ab1791) or γH2A.X (Abcam ab2893). The tubes were sealed with

parafilm and kept at 4ºC rotating overnight. Protein G magnetic beads were added to the tube with a cut pipette tip (30 µL/tube). Tubes were sealed in parafilm and incubated with tumbling (4ºC, 1-2 h). The beads were separated on a magnetic rack and the liquid was removed. 750 µL wash buffer one (0.1% SDS, 1% triton X-100, 2 mM EDTA, 20 mM tris pH 8 and 150 mM NaCl) was added per tube and the tubes were inverted 25 times to disperse the beads. The beads were separated on the magnetic rack as before, 750 µL wash buffer two was added (0.1% SDS, 1% triton X-100, 2 mM EDTA, 20 mM tris pH 8 and 500 mM NaCl) and the tubes were inverted 25 times to disperse the beads as before. This was repeated with wash buffer three (1 mM EDTA, 10 mM tris pH 8, 1% NP-40, 1% sodium deoxycholate and 0.25 M LiCl) and wash buffer four (10 mM tris pH 8 and 1 mM EDTA). A short centrifuge was performed to remove the remaining liquid and the total inputs were thawed. Elution buffer (100 mM $NaHCO_3$ and 1% SDS) was added, at 150 µL per sample or 100 µL for inputs. RNaseA (20 µg) was added to each sample and they were incubated in the thermomixer (1000 rpm, 65ºC, overnight). The samples were removed from the thermocycler and treated with Proteinase K, as before, then incubated at 42ºC for one hour.

2.6.2.6 Phenol-chloroform extraction

The extraction was performed in a fume hood. 150 µL MPW followed by 300 µL phenol-chloroform was added to each tube, the tubes were shaken and spun down (16,000 *g*, 5 min, RT). The top, aqueous phase was transferred to a new tube (DNA lo bind), 300 µL MPW was added, the tubes were shaken and spun down as before. Chloroform (550 µL) was added to each tube, and the tubes were shaken and spun down again. The top phase was transferred to a new tube (DNA lo bind), 300 µL MPW was added, shaken and spun down as before. The top phase was transferred to a new tube with 2 µL glycogen and 0.5 M NaCl. One volume of isopropanol was slowly added to the top of the tube and the tubes were transferred to sealed Eppendorf® racks where the DNA precipitated (overnight, -20ºC). The tubes were spun (16,000 *g*, 1 h, 4ºC), the supernatant was removed and 150 µL 70% ethanol was added. The tubes were centrifuged (16000 *g*, 2 min, 4ºC), the supernatant was removed, and this was repeated. All ethanol was removed, the pellet was dried (5 min, RT) and resuspended in 10 µL EB (Qiagen).

2.6.2.7 Qubit fluorometric dsDNA quantification (Thermo Fisher)

A working solution was made (199 µL dsDNS HS buffer, 1 µL Qubit dsDNA solution) for the number of samples to be tested. The standards were then prepared in Qubit dsDNA tubes (190 µL working solution with 10 µL standard one or two) with the samples (199 µL working

solution, 1 µL sample) and these were vortexed and left at room temperature for three minutes. The standards were then read in the machine, followed by the samples which then had their concentration recorded. DNA libraries were then prepared.

2.6.2.8 DNA library preparation

DNA libraries were made using the NEBNext® Ultra™ II library prep kit (E7645) using 1 ng of starting DNA. Libraries were prepared as stated in the kit instruction manual using eleven cycles of PCR enrichment. The libraries were checked using the bioanalyzer, quantified using the KAPA system and sent for sequencing as explained in 2.4.1.4-6.

2.7 Data analysis

2.7.1 Analysis of sequencing data quality

To assess the quality of the data sequences were analysed with FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The FastQC report was checked for the per base sequence quality, per tile sequence quality, per sequence quality score, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences, adapter content and K-mer content. If these were within the acceptable range for either RNAseq or ChIPseq data then the data was analysed further by the PNAC Bioinformatician at the MRC LMB, Alastair Crisp. Data was trimmed for adaptor sequences with a minimum quality threshold of 30 using TrimGalore v0.4.4 (https://github.com/FelixKrueger/TrimGalore).

2.7.2 Performing differential expression analysis

Trimmed reads were aligned to the Human genome version GRCh38.87 using TopHat v2.1.0 (https://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-4-r36), reads were quantified per genomic region and differential expression was calculated using Cufflinks v2.2.1 (https://www.nature.com/articles/nbt.2450). Separately, a principle component analysis plot (PCA) for each experiment was produced using the TopHat alignment and DESeq2 v1.18.1 (Love et al., 2014) and the similarity of each triplicate checked before further analysis. CuffDiff was used to calculate differential expression analyses, using a student's t test and assuming a normal distribution T=(E[log(y)])/(Var[log(y)) where E is expression, Var is variance and y is the ratio of the normalized counts between the two conditions. The FDR was set to 0.05. Further analysis will compare the results of the Cufflinks and DESeq2 pipelines.

2.7.3 Analysing sequencing data

All further analysis was performed by myself. Triplicate samples for each experiment were loaded into the PEAT analysis tool (https://github.com/lmb-seq/PEAT), generated by Alastair Crisp and Paula Freire Pritchett at the MRC LMB, for further analysis.

Data was checked for differentially expressed genes and read counts, GO terms (gProfile) and TF binding sites (TRANSFAC geneXplain), and plots could be generated. Probability testing of enrichment analysis was performed using Benjamini Hochberg multiple testing.

Significantly differentially expressed genes were taken to be transcripts that had a FPKM of $\geq$ 1 in each sample and a $\log_2$ fold change of one between samples.

Chapter 3. Results I: Setting up an *in vitro* system to study differentiation

3.1 Mouse embryonic stem cells

On initiating the project, I elected to use mouse embryonic stem cells (mESCs) to explore two different changes of state in cell culture. The two mouse ES cell lines used were the well-established male E14 and female PGK12.1 lines (Hooper et al., 1987; Norris et al., 1994). Both are feeder-free, leukaemia inhibitory factor (LIF) dependent cell lines that can be cultured without feeders on gelatin-coated dishes.

3.1.1 Culturing mESCs: the 'gold standard'

Mouse ES cells can be maintained in either of the serum and LIF or 2i and LIF states although their gene expression profiles, epigenetic marks and morphology are remarkably different between the two. The interconversion between these two states is possible by switching the media in just a few passages. The transition between these states was monitored throughout the chapter, along with the transition to epiblast-like cells.

3.1.2 Transition from mES cells to epiblast-like stem cells

Prior to differentiating the mESCs, cells were grown in LIF and 2i containing culture medium for a few passages. Transferring the cells from serum to 2i was straightforward but the passaging of cells required optimisation to ensure the cells re-adhered after each passage. This could be helped by using specific culture dishes (Corning™ Costar™). Once mES cells were adapted to this state, they were differentiated to epiblast-like cells (EpiLC); this transition mimics the transition from the preimplantation embryo, mES cells, to the post-implantation embryo, EpiSCs. In this process, the cells transition from naïve pluripotency to primed pluripotency whereby one X-chromosome is inactivated in female cells and the cells become dependent upon Activin A and FGF2 (Nichols and Smith, 2009). Cells were cultured on fibronectin and N2B27 medium was supplemented with Activin A and FGF2 for at least three days prior to assessing the cells for loss of mES cell markers and gain of EpiLC gene expression: this transition took around five to six cell cycles and therefore would allow a replication-dependent phenomenon to be shown. The transition between the two states was reversible within the first 24 hours, after which the cells lost the mES cell state markers and were committed to differentiation. Preliminary experiments showed that the viability of the cells, as

examined by the cell counter, was lower than expected. The number of cells plated before initiating differentiation was increased to over $1.6x10^5$ cells per well of a six-well plate and this increased the cell viability to above 90 percent from 70 percent originally.

3.1.3 Analysing differentiation efficiency of mESCs using qPCR

Gene expression changes occurring during this transition were monitored using quantitative PCR (qPCR) which allows quantification of the mRNA levels of specific transcripts in the bulk population of cells over this period. Although the core trio of pluripotency genes, *Pou5f1, Sox2* and *Nanog*, remain expressed at relatively high levels from mESCs to EpiLCs, there are several genes that become downregulated on differentiation to the epiblast-like state (Guo et al., 2009). These genes include *Zfp42, Esrrb, Dppa3* and *Klf4*; they are key transcripts to probe using qPCR to validate the system. As well as using the genes that will turn off over the course of a successful differentiation, it is important to look at genes that are switched on during this process. This gives a deeper knowledge into the direction of the differentiation, suggesting whether the cells simply lose pluripotency or whether they become specified down the EpiLC pathway. There are a number of genes that are upregulated at this transition including *Nodal*, brachury *(T)*, *Fgf5* and goosecoid *(Gsc)* (Guo et al., 2009).

qPCR primers were designed for a selection of genes that would be differentially regulated during the differentiation: the primers needed to produce a 100-180 bp sequence when amplified. Where possible, this was designed over an intron such that only cDNA and not gDNA would be amplified. Housekeeping control genes were also picked to normalise the results to, however considering the nature of the differentiation some common 'control' genes were likely to alter significantly over this process.

Preliminary results showed that the differentiation protocol was efficient: viewing cells down the microscope showed that the morphology was altered dramatically from forming small round colonies to producing individual cells with a flatter nature. RNA was collected in the mES and EpiL cell states and converted to cDNA before being quantified using qPCR (Figure 9). qPCR data showed that gene expression patterns had robustly changed in the direction expected during the differentiation for all genes apart from *Nodal* which showed no difference between the two states. However, one clear problem was that the two normalising genes initially being used, *β-actin* and *Gapdh*, gave slightly different relative readouts, suggesting that these transcripts may also change expression during differentiation (Figure 9a and b). *Gapdh* was used to normalise

gene expression in all further experiments due to gene expression remaining most constant in the literature during mES cell differentiation (Murphy and Polak, 2002).

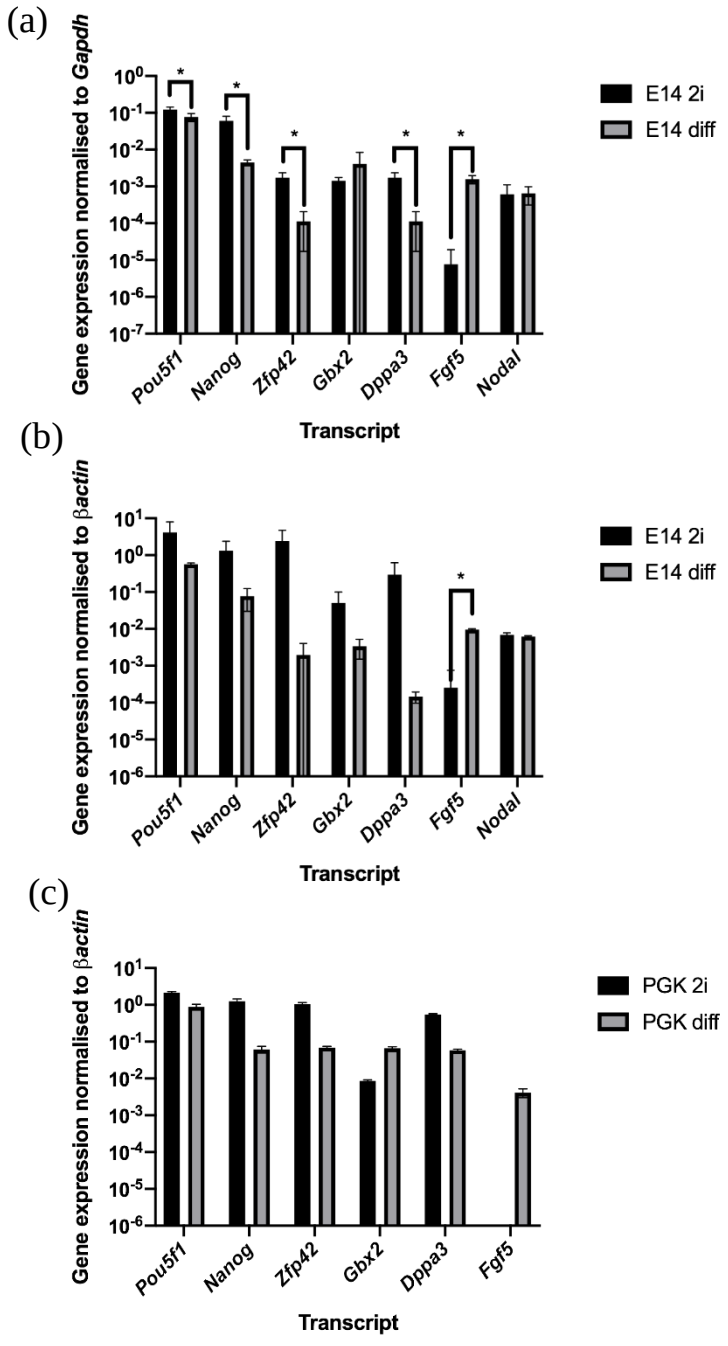


Figure 9. qPCR data showing the differentiation of mESCs to EpiLCs. (a) E14 mES cells were cultured in 2i medium (E14 2i) before being differentiated to EpiLCs (E14 EpiLC), RNA was extracted in each stage and qPCR was performed. Gene expression was normalized to *Gapdh*. Mean±SD are plotted of three technical replicated of n=4 biological replicates. Multiple t tests used to calculated $p<0.05$ marked with *. (b) as (a) but normalising to *βactin* n=2. (c) using PGK12.1 cells n=1 biological replicates, error bars show technical replicates in this instance.

The PGK12.1 mES cells did not differentiate as efficiently as the E14 cells, as seen by the level of cell death and qPCR data (Figure 9c). Most experiments, especially those involving differentiation, were only performed in E14 cells in order to give the most reliable data. Most published experiments use only male cell lines and this may be in part due to the slower differentiation and altered gene expression profile of female cells containing two X-chromosomes, due to the process of X-inactivation (Rastan and Robertson, 1985; Schulz et al., 2014).

3.1.4 DNA replication impediments during differentiation

Having set up a system to study embryonic differentiation, the aim of the project was to assess whether problems during DNA replication could alter the course of differentiation. While there are numerous ways to cause replication impediment formation, I began by looking at G-quadruplexes (G4s). There were two means by which I planned to do this: culturing cells with G4-binding ligands which are able to bind to transient G4s and stabilise them in the genome (De Cian et al., 2007; Guilbaud et al., 2017; Li et al., 1996; Rodriguez et al., 2012) and by producing genetic knockouts of enzymes known to be involved in processing G4s, likely increasing the persistence of the structures at the replication fork: *Rev1*$^{-/-}$, *Brip1*$^{-/-}$ *(Fancj)* and *Primpol*$^{-/-}$ (Sarkies et al., 2012; Schiavone et al., 2014, 2016).

3.1.4.1 G-quadruplex-binding ligands

Initially, two G4-binding ligands were chosen to assess whether stabilising G4 secondary structures in the genome could alter the efficiency of differentiation: Pyridostatin (PDS) and PDC12. PDS was tested because it is frequently used in G4 experiments, has been well documented in the literature and has a high affinity for a broad range of G4s (Rodriguez et al., 2008). However, it is also thought to induce DNA damage and therefore its effects on the cells may be broader than simply stabilising G4s (Rodriguez et al., 2012). PDC12 was used as it had been developed by the Balasubrumanian group as a PDS derivative and its *in vivo* G4 stabilisation activity had been analysed by our lab (Guilbaud et al., 2017). It was shown to induce less toxicity and DNA damage in cells compared to previous G4 ligands, allowing them to be cultured for longer periods of time (Guilbaud et al., 2017; Rodriguez et al., 2012). However, as discussed later, there were numerous issues related to using this drug in mES cells, namely that it required much higher concentrations to induce an effect and was highly insoluble in DMSO, which in turn could lead to adverse effects on differentiation. I hypothesised that using these two ligands in parallel would allow us to gain insight into which effects may be specific to G4s. However, they may well bind to structurally different G4 structures: this is discussed in greater detail throughout the thesis.

A range of concentrations of both drugs were tested on the cells to assess the toxicity to the mESCs and this was compared to the literature (Guilbaud et al., 2017). PDC12 and DMSO were added to E14 and PGK12.1 cells cultured in the 2i state to analyse how these treatments affected the viability and doubling time of the cells at a range of concentrations after six days: 0.3125 to 40 μM (Figure 10). This shows that the cells exhibited similar viability and doubling times across the range of DMSO and PDC12 concentrations tested in the 2i state. PDS

concentrations were kept between 0.5 and 4 μM for all experiments but there was significant batch to batch variation in the drug, such that inducing the same effect often required slightly varying concentrations. The concentrations of PDC12 were initially kept between 10 and 20 μM, which had been shown to exert a significant effect on the endogenous G4 in the BU-1 locus, but were then increased up to 80 μM as minimal effects were induced. These concentrations meant the addition of large volumes of DMSO which, despite having little effect in DT40 cells (Guilbaud et al., 2017), were likely to affect differentiation to an extent itself; this will be discussed in greater detail with reference to the hiPS cell differentiation. Another issue that became noticeable with the PDC12 ligand was that it formed crystals ≥ 10 μM in all conditions tested. These crystals were large, and it was clear that the true concentration of drug in the media was significantly lower due to this crystallisation. The crystals could be removed by filtration but this did not overcome the concentration reduction. Although the following experiments were completed with both PDS and PDC12 in parallel, only the results with PDS will be discussed further due to the inconsistencies with PDC12 at high concentrations, and therefore large DMSO volumes, required to give a G4 specific effect. Another G4 ligand, NMM was used during the project when it was clear that the PDC12 was unreliable, and this is discussed throughout the chapter (Nicoludis et al., 2012). This ligand was used at concentrations between 0.5 and 4 μM, similarly to PDS, at a level that did not induce significant toxicity.

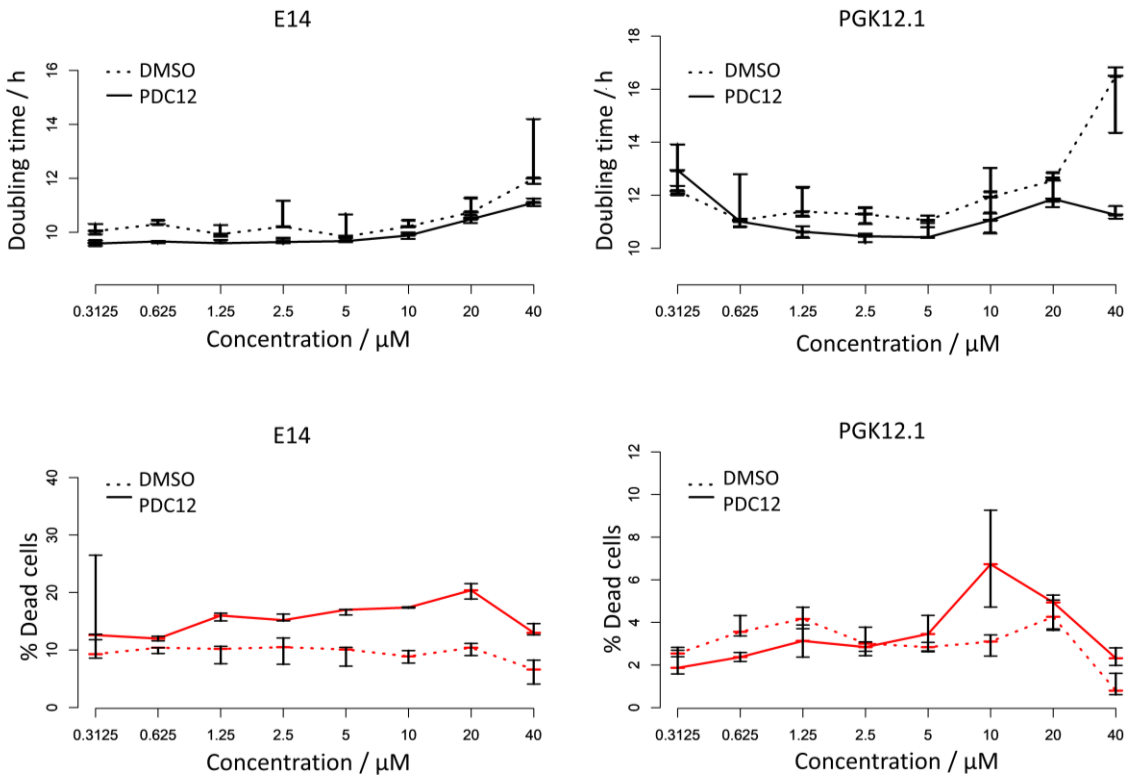


Figure 10. E14 and PGK12.1 mES cells were cultured in 2i medium for six days with either DMSO as a vehicle control or the G4 ligand PDC12 at concentrations from 0.312 to 40 µM. The number of live cells were counted using PI staining and this allowed the doubling time and the percentage of dead cells to be quantified.

The setup of different experiments with the G4 ligands was able to address key questions (Figure 11):

1. Can maintenance of the ground state be perturbed by G4 ligands and can this alter differentiation?
2. Can perturbing the ground state and differentiation with G4 ligands alter the differentiation process?
3. Can G4 ligands alter the differentiation process directly?

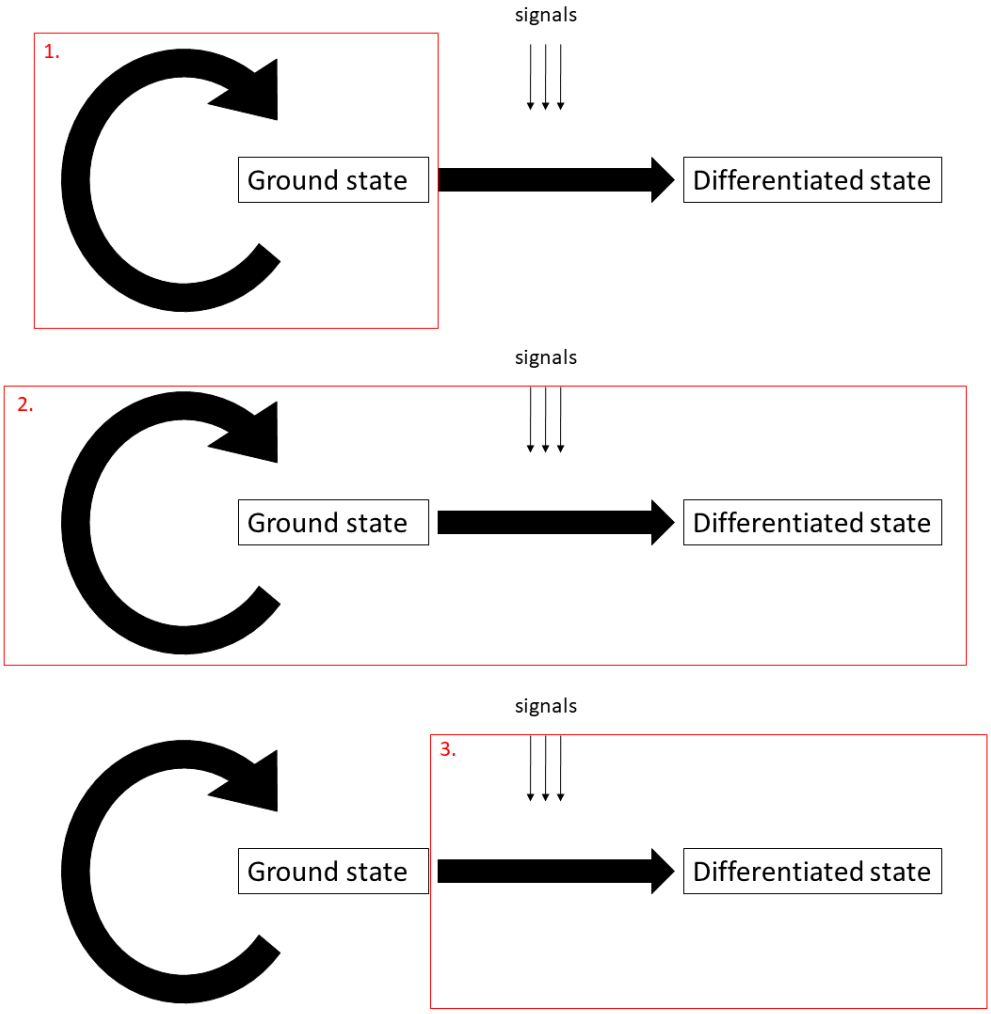


Figure 11. Schematic to show different targeting of the ligands in the undifferentiated state and during differentiation. (1) assessing whether culturing with ligands in the ground state can impair the ground state maintenance and the propensity to differentiate (2) assessing whether ligands in the ground state and during differentiation, impact differentiation (3) addition of ligands at the onset of differentiation to understand whether this can perturb differentiation.

Can maintenance of the ground state be perturbed by G4 ligands?

The initial experiment addressed whether maintaining the cells in the presence of G4 ligands in the 2i state could alter the expression of pluripotency markers. E14 and PGK12.1 cells cultured in the presence of ligands for three weeks were collected and RNA was extracted for qPCR analysis (Figure 12). It was clear that PDS treatment in both E14 and PGK12.1 cells, as well as NMM in the E14 cells, resulted in a decrease in pluripotency marker genes including *Gbx2* and *Zfp42*. This suggested that either these drugs were able to alter the expression of pluripotency

genes in the ground state and this could alter the viability of the cells or they could alter the viability which could change the expression of pluripotency factors. However, whether these ligands were inducing a replication-dependent G4 effect could not be ascertained with this data.

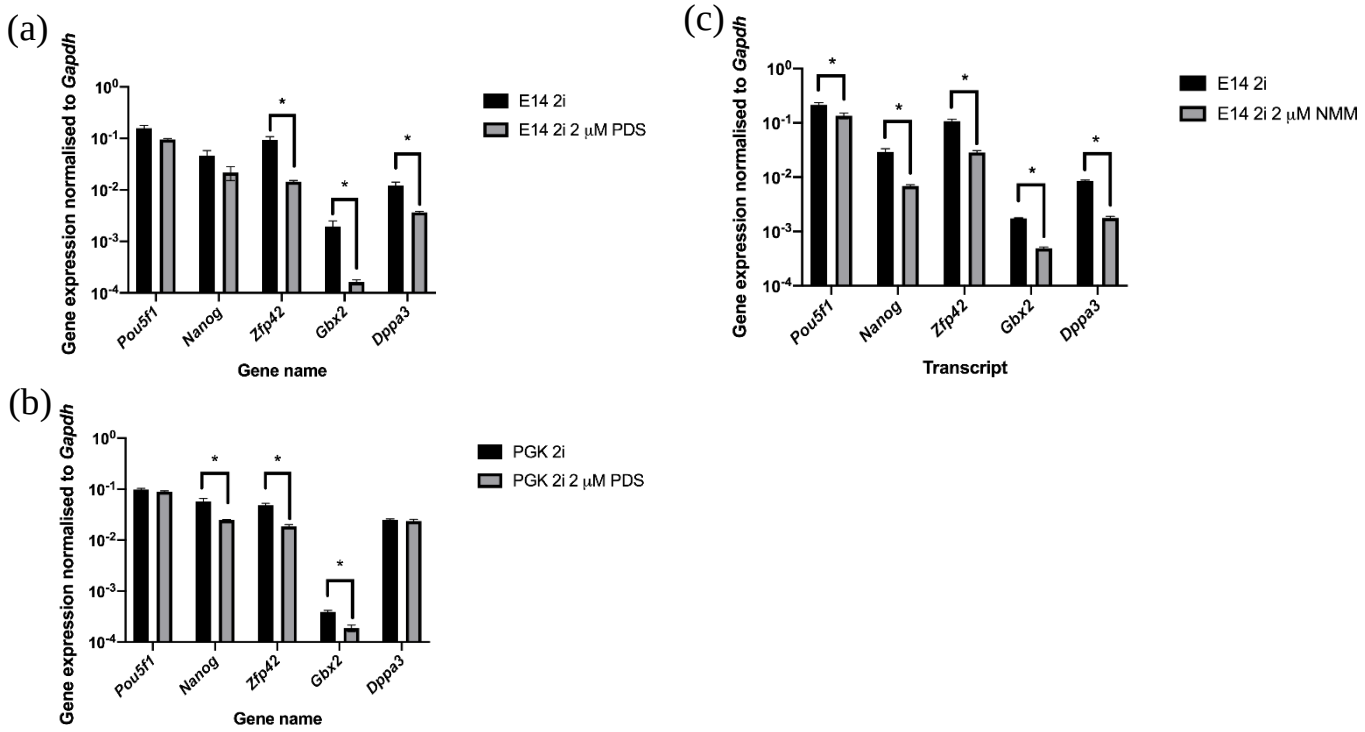


Figure 12. qPCR data normalised to *Gapdh* for pluripotency genes in mESCs cultured in 2i medium with and without the addition of G4 ligands for three weeks (a) E14 cells with 2 μM PDS (b) PGK12.1 cells with 2 μM PDS (c) E14 cells cultured with 2 μM NMM. Mean is plotted with error bars showing the standard deviation of technical triplicates and * denotes significance $p<0.05$ using multiple t testing.

Can perturbing the ground state with G4 ligands alter the differentiation process?

The next experiment aimed to probe whether altering the ground state using G4 ligands could change the efficiency of differentiation to EpiLCs. E14 mESCs were cultured with PDS for three weeks and then washed to remove the drug prior to differentiating the cells without the ligand. Cells were collected eight days after the induction of differentiation to EpiLCs, with and without PDS, and RNA was extracted for qPCR (Figure 13). This initial result suggested that culturing the cells in the ground state with the ligand altered the differentiation efficiency very little, however this experiment was performed using a low concentration of PDS, 0.5 μM, due to toxicity at high concentrations using this batch.

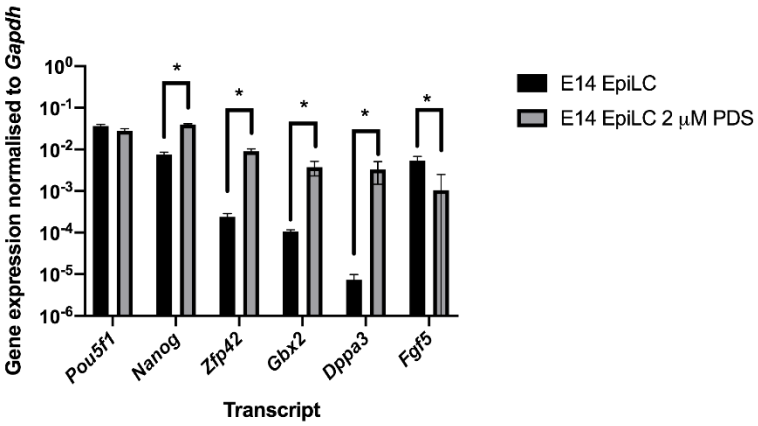


Figure 13. qPCR data for pluripotency and EpiLC transcripts in E14 mESCs cultured with or without 0.5 μM PDS in the 2i state for three weeks before washing out the ligand and differentiating without PDS. Error bars are the standard deviation of technical triplicates. * denotes significance of $p<0.05$ using multiple t testing.

Can G4 ligands in the ground state and during differentiation alter the differentiation process?

The next experiment cultured the cells with ligand in the mESC state and then also differentiated the cells in the presence of the ligand. Cells were cultured with the ligand for three weeks and then differentiated in the presence of 2 μM PDS, RNA was collected after four days and qPCR was performed (Figure 14). The presence of PDS in the E14 cells did not significantly prevent upregulation of markers of EpiLCs. However, these treated cells did not show a significant decrease in mESC gene expression compared to untreated cells including *Nanog, Zfp42, Gbx2* and *Dppa3*. This suggested that these cells may be remaining in a more pluripotent state and not differentiating. Together these experiments showed that G4-binding ligands were able to alter the pluripotent state and the efficiency of the differentiation process. However, since the qPCR experiments were performed only once the significance of this results cannot be confirmed using this method. The alkaline phosphatase assay could have also been used to analyse the cell state of the mES cells to confirm adequate differentiation.

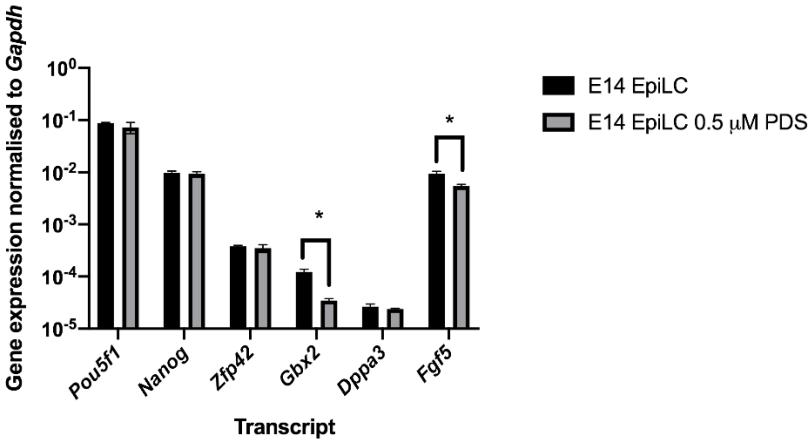


Figure 14. qPCR data normalised to *Gapdh* showing E14 cells grown in the undifferentiated state with 2 μM PDS for three weeks and then differentiated in the presence of the drug for four days. Error bars are the standard deviation of technical triplicates. * denotes significance of $p<0.05$ using multiple t testing.

3.1.5 Single-cell analysis of differentiation efficiency

While performing these qPCR experiments it was clear that there was not going to be a clear-cut answer to the initial question: G4-binding ligands would not fully prevent differentiation, but they were likely to be altering the course to an extent. This approach for looking at the bulk population, while a good way for checking that the differentiation protocol was working successfully, was not going to allow us to view the results on a single-cell basis. If the changes being seen were replication-dependent, stochastic changes that were able to alter the epigenetic state of a cell, as previously described in the lab, then we would expect to see fluctuations in gene expression on an individual cell level (Sarkies et al., 2012). The simplest way to assess the differentiation in single cells would be using flow cytometry which is able to sort the fluorescence of each cell individually, and therefore monitor the gene expression at a cellular level. If these experiments were to be repeated, using single-cell RNA sequencing would have also been a very interesting approach, however at the point when these experiments were performed, we did not have the resources or bioinformatics power to perform this approach.

3.1.6 Permeabilised flow cytometry to monitor differentiation

Initially, permeabilised flow cytometry was performed using antibodies against REX1 as a marker of pluripotency, and therefore the mESC state, and against DNMT3A as a marker of the epiblast state. Cells maintained in the 2i state, as well as those in the EpiLC state, were collected and fixed, they were then permeabilised, stained for REX1 or DNMT3A protein expression and analysed using flow cytometry. The REX1 antibody did show a decrease in protein expression in the EpiLCs but there was not a complete separation between the two states (Figure 15a). The antibody against DNMT3A did show an increase in protein expression throughout the differentiation process as would be expected in a successful differentiation, but again there was not a clear separation between the cells cultured in 2i and the differentiated state, shown in Figure 15b, and this was not consistent between experiments. To give an insight into whether it was a problem with the differentiation or the antibody, *Dnmt3a* knockout mES cells known not to express protein (a kind gift from Gerry Crossan at the MRC LMB) were used to validate the antibody (Figure 15c). Despite the cells not being expected to express the protein, there was DNMT3A protein signal higher than the unstained secondary control, which should not have occurred. There was a difference between the knockout cells and differentiated cells but the difference was insufficient to see small changes: this antibody was not ideal for permeabilised flow cytometry and may well have been binding non-specifically. A number of optimisations were trialled but with no improvement to the results. Other research groups have tried to

optimise permeabilised flow cytometry for this cell state transition but have not found any successful antibodies for this. This approach was taken no further.

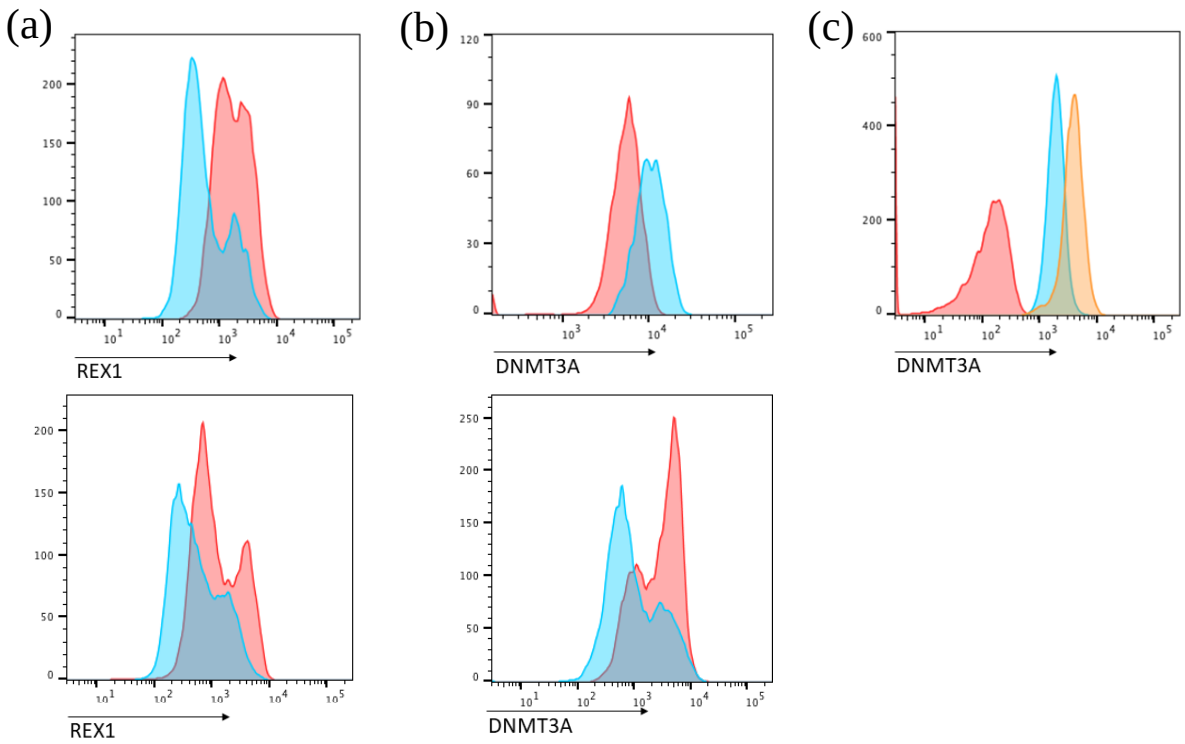


Figure 15. Permeabilised flow cytometry data in E14 mESCs, data shown as histograms, repeats of different experiments are shown below giving the variability of this staining (a) staining for REX1 as a marker of the 2i state in cells grown in 2i (red) and EpiLCs (blue) (b) staining for DNMT3A as a marker of the EpiSC state in cells grown in 2i (red) and EpiLCs (blue) (c) staining for DNMT3A in EpiLCs (orange), *Dnmt3a-/-* EpiLCs (blue) and unstained cells (red).

3.1.7 *Gfp* tagged genes to monitor pluripotency

Another way to monitor expression on a single-cell basis was to use cells expressing fluorescently tagged pluripotency marker genes, such that they should be expressing GFP in mES cells but not in EpiLCs. The first cell line used was acquired from Austin Smith's group: the pluripotency gene *Zfp42* had been tagged with *Gfp* to monitor its expression during differentiation (Wray et al., 2011). The GFP used was a destabilised version (GFPd2) inserted into the *Zfp42* locus of $\beta cat^{fl/-}$ E14 embryonic stem cells, using a less stable version meant that the protein did not remain in the cell for a long time after being synthesised. This was particularly useful during differentiation, allowing a decrease in expression to be seen quickly and giving an accurate readout for the gene expression of *Zfp42*. The other GFP-tagged cell line was procured from Gerry Crossan, whereby the ES cells were derived from transgenic mice as described previously (Czechanski et al., 2014). The transgene was made using the BAC9 construct in which *Dppa3* was tagged with an enhanced version of GFP

(e*Gfp)* before the 3'UTR and it was shown to behave in a similar manner to the endogenous gene (Payer et al., 2006).

Preliminary experiments in both GFP marker cell lines showed a clear difference in expression when comparing cells cultured in 2i compared to differentiated cells, shown in Figure 16. However, there was a larger divide between the positive and negative populations in the *Zfp42-Gfp* cell line, possibly due to the GFPd2 protein variant being expressed. I decided to continue the following experiments in the E14 *Zfp42-Gfp* cell line because this was in the same mouse ESC background as my previous experiments and the destabilised GFP gave a clearer separation between the two fluorescence profiles. E14 *Zfp42-Gfp* cells were cultured with G4-binding ligands in the pluripotent mESC state; as previously, PDS and NMM were used during these experiments.

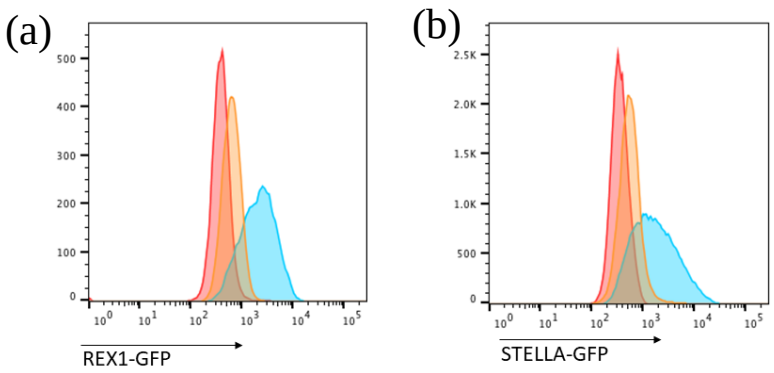


Figure 16. Flow cytometry analysis of mESCs with pluripotency GFP reporters in the 2i (blue) and EpiLC states (orange) with control cells not expressing GFP (red) (a) *Zfp42-Gfp* (b) *Dppa3-Gfp*.

Cells were cultured in the 2i state in the presence of either G4 ligands for three weeks and GFP-fluorescence was monitored by flow cytometry: this was compared to untreated *Zfp42-Gfp* cells (Figure 17a and b). This shows that there was no difference in the GFP fluorescence in the 2i state of the cells grown with or without the drugs, suggesting that REX1 was not altered in the undifferentiated 2i state. Previous qPCR data suggested that there were small changes in the *Zfp42* RNA in PDS and NMM treated cells in the 2i state, this will be discussed later.

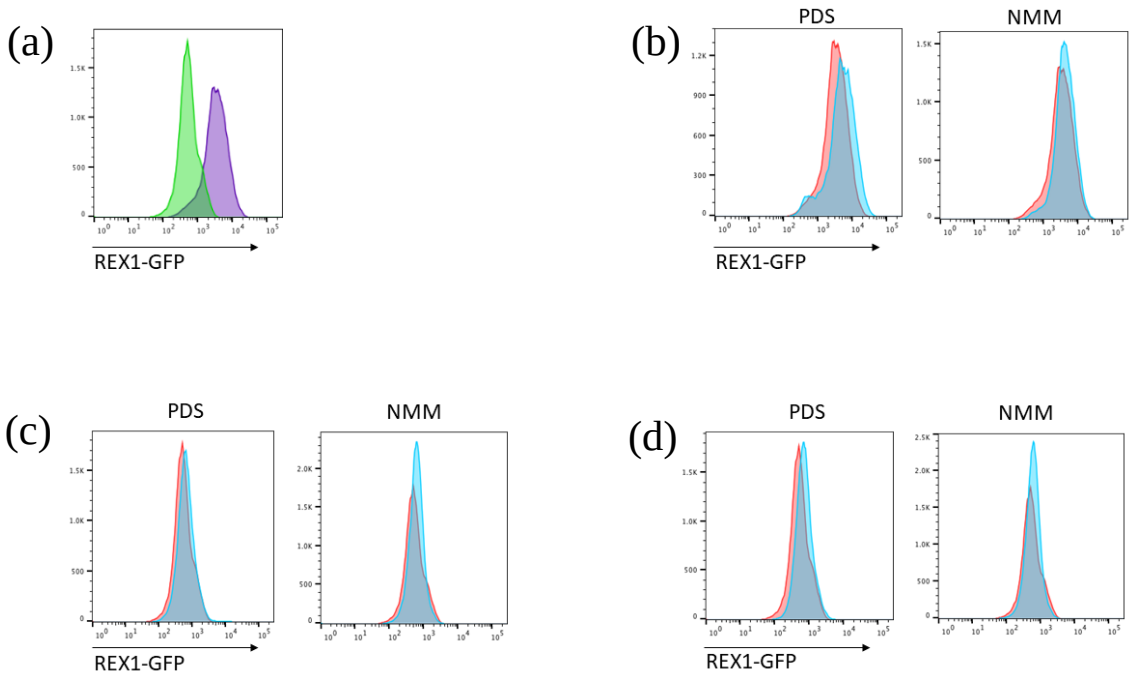


Figure 17. *Zfp42-Gfp* E14 mESCs were monitored for GFP in the 2i or EpiLC states with or without 2 μM PDS or NMM (a) untreated cells in 2i (purple) and EpiLC state (green) (b) cells cultured in 2i with PDS (left panel) or NMM (right) for three weeks, untreated (red) and treated (blue) (c) cells were grown on ligands in 2i for three weeks and differentiated in the presence of drugs for three days as (b) (d) cells were differentiated with ligands without being grown on them in the 2i state as (b).

The cells were also maintained in 2i with the drugs present and then differentiated with the drug to see whether this could alter the proportion of cells that lost GFP expression (Figure 17c). This experiment was repeated with the cells grown without drugs in the 2i state and then differentiated in the presence of the drug for 96 hours (Figure 17d). In each of the differentiated conditions the majority of the cells stopped expressing REX1-GFP and while this suggests that they may all have been differentiating well, REX1 was marking a loss of pluripotency rather than an end point of differentiation. While this experiment was only performed once and therefore the significance cannot be determined, the preliminary results suggested that there was not a difference. Therefore, a lack of GFP merely suggests that the cells have lost pluripotency. Interestingly, the mESCs were able to tolerate the ligands for much longer than the EpiLCs and could be maintained on them in both the 2i and serum conditions for up to three weeks. Whether the ligands were inducing genetic or epigenetic changes was unknown and this difference is discussed in greater detail in Results III with reference to the hiPSCs which showed the same trend.

As this assay only gives a read-out for the loss of pluripotency and not the efficiency of differentiation, a GFP-tagged EpiSC marker gene would have been the most useful for analysis.

However, the pluripotency state of the cells helped to address the first question of whether the ligands can perturb the ground state. Interestingly, in the mESC system, cells which lose pluripotency when being maintained on 2i no longer adhere onto the culture dish and instead float to the surface of the medium. This was particularly obvious after passaging the cells but also happened throughout their culture; this was visible as 'floating cells' in the media, either as individual cells or clumps. The cells had lost pluripotency but may have remained alive, and therefore the G4 ligands may been altering the proportion of cells losing pluripotency but this was hard to confirm using flow cytometry of the cells which have remained adherent: to the eye, there were many more 'floating cells' in the drug-treated cells. Therefore, the REX1-GFP experiment may have been missing the cells that lose adherence and introducing bias.

In order to investigate whether these cells that lost adherence were altering the proportion of REX1-GFP positive cells in the 2i state, the cell culture medium containing cells in suspension was analysed using flow cytometry. Some of the cells which lost adherence would be dead cells, therefore propidium iodide (PI) staining was performed in order to discount dead cells. By this method, screening for GFP-, PI- cells selects for the population of non-pluripotent, alive cells. Pluripotent cells maintained in the 2i state could be grown in the presence of drugs for around three weeks but on every passage there were fewer cells that re-adhered to the culture dish suggesting a loss in pluripotency over time. To test this hypothesis, the supernatant of cells grown in this manner was collected and analysed using flow cytometry for GFP- PI- cells. While there were fewer cells in the supernatant to monitor than adherent cells, it was possible to pool the supernatant from a number of wells to gain sufficient quantity for analysis. This experiment was performed in the presence PDS or NMM and the viable cell population in the supernatant was analysed and compared to the viable adherent cell population (Figure 18). These results suggest that while the adherent cells in each condition were uniformly expressing GFP, the supernatant cells treated with the drugs that were still alive did not express GFP and therefore had lost pluripotency, suggesting that this was why the cells had lost aherence. This was very clear in the PDS treated cells. In the untreated cells there was a large range of cells expressing and not expressing GFP, implying that lack of pluripotency was not the main reason for these cells to lose adherence but clearly did occur. However, when the cells were cultured for two weeks, this difference was no longer evident, perhaps because the toxicity of the ligands induced greater levels of cell death after culturing for this length of time. This result suggests that maintaining the cells in G4 ligand in the 2i state was able to alter the expression state of the cells, causing them to lose pluripotency and therefore stop adhering. The cells collected for RNA extraction may have contained greater number of 'floating cells' compared to those

collected for flow cytometry due to the different methods which may explain the difference in the *Zfp42*/REX1 levels between these two techniques. Alternatively, this could be explained by differences in RNA and protein expression not changing at comparable rates.

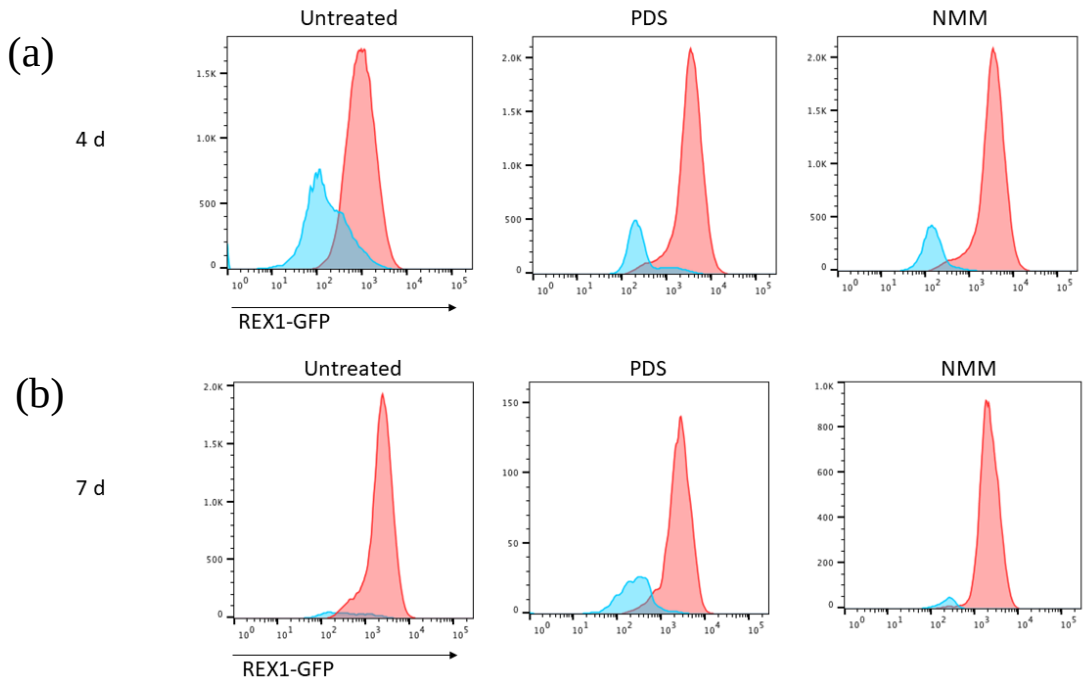


Figure 18. Flow cytometry data gating for PI- live cells in the supernatant (blue) compared to attached cells (red) in mESCs cultured in 2i. Untreated cells (left panel), 2 µM PDS treated (middle) and 2 µM NMM treated (right) (a) four days in the ligand (b) seven days.

In the mouse E14 ES cells, genetic knockouts of *Rev1*, *Brip1* and *Primpol* were performed using a CRISPR/Cas9 strategy but verification was not fully completed before making the decision to move onto using the human iPSC differentiation strategy. This will not be discussed in further detail.

While this system gave some initial results to suggest that stabilising G4 structures may alter mouse embryonic development, it was difficult to monitor the efficiency of this transition on a single-cell basis and the efficiency was not high. The qPCR results showed that treating with G4 ligands decreased the expression of pluripotency markers in the 2i state, and the flow cytometry data using GFP markers showed that cells that lost pluripotency ended up losing adherence, and this happened to a greater extent when treated with PDS or NMM ligands. This transition from mES cells to EpiLCs is considered to be more of a ‘state change’ than a ‘differentiation’ and despite being useful for preliminary experiments it was not particularly useful to test the initial hypothesis. Therefore, with significant cell culture teaching from the Vallier group, I moved onto hiPS cells and the differentiation to definitive endoderm.

3.2 Human induced pluripotent stem cells

In order to focus on a more definitive differentiation protocol, a human iPSC line was chosen to differentiate, the BOBSC cell line (Yusa et al., 2011). This iPSC line was chosen as it had been employed by the Sanger Centre COMSIG project which therefore provided access to a range of genetic knockout cell lines. Cells can be maintained in the pluripotent state in a similar manner to the mESCs, with the main difference being that the human cells do not tolerate being maintained as single cells and are therefore kept as small colonies throughout the passaging process. These cells are considered to have a high efficiency of differentiation, highly valuable for these experiments.

3.2.1 Differentiation to definitive endoderm

I considered a number of differentiation systems with the main requirements being:

1. A high enough efficiency of differentiation that relatively small deviations in the efficiency could be analysed.
2. At least three cell cycles would occur in the normal process to allow for a replication-dependent epigenetic change to be realised.
3. A well described system whereby the major pathways were known.
4. The possibility to monitor key gene expression changes on a single-cell basis by flow cytometry.

Taking these factors into account, the differentiation of iPS cells to definitive endoderm was chosen. This occurs over 72 hours, during which the cells undergo around five cell cycles, and can achieve up to 90 percent efficiency, meaning that this system matched my criteria. The specific protocol used was provided by Rodrigo Grandy in Ludovic Vallier's group in Cambridge, adapted from (Teo et al., 2011).

These iPS cells were cultured in the pluripotent undifferentiated state on vitronectin coated culture plates with Essential 8™ medium; cells were maintained in the undifferentiated state and split onto a new plate one day before initiating the differentiation. To initiate the differentiation to definitive endoderm, the cell media was replaced with CDM-PVA media and insulin, the WNT pathway activator CHIR99021, the PI3K inhibitor LY294002 and the recombinant proteins Activin A, FGF2 and BMP4. After 24 hours the cells were already morphologically distinguishable from undifferentiated cells as the cells became more individual in nature and the cell boundaries were visible at the edges of colonies. However, I noticed a lot of cell death during this part of the differentiation as seen by the number of cells which lost adherence, but this is a qualitative observation. On the second day the WNT activator was

removed from the differentiation media and the cells begun to express WNT pathway inhibitors. At 48 hours the cells went through the epithelial mesenchymal transition (EMT) and the colonies became less visible. The BMP4 and LY294002 were removed at 48 hours and only Activin A and FGF2 were maintained in the media for the final 24 hours. At 72 hours the cells were harvested to assess their efficiency of differentiation.

This system was used to answer a number of questions that were not fully addressed over the mES cell transition.

1. Do G4 ligands perturb the ground state or alter the fate of definitive endoderm differentiation?
2. Can genetic knockouts of enzymes known to be involved in processing G4s change gene expression in the pluripotent state or change the efficiency of differentiation?
3. Do other replication impediments act in a similar manner to G4s in the undifferentiated state and during differentiation?

3.2.2 Monitoring the efficiency of definitive endoderm differentiation

The differentiation was monitored using methods including permeabilised flow cytometry and quantitative PCR, as previously discussed in the mouse system. Permeabilised flow cytometry using two antibodies to proteins expressed during different stages of the definitive endoderm differentiation gives a clear, quantitative read-out of the efficiency on an individual cell basis; this was missing from the mouse system. Eomesodermin (EOMES) is expressed from around eight hours into differentiation and therefore cells harvested at 24 hours express high levels of EOMES compared to the undifferentiated cells which do not express EOMES. The expression of EOMES decreases throughout differentiation but some protein remains at 72 hours. SOX17, a protein belonging to the SRY-box family, is considered to be a marker of definitive endoderm specification and the proportion of cells expressing this protein at 72 hours can be considered a read-out of the efficiency of a particular differentiation experiment (Teo et al., 2011).

The cells were collected at the desired timepoint, fixed and then stored until performing flow cytometry. To perform flow cytometry, the cells were thawed, permeabilised, and stained with EOMES and SOX17 antibodies. The addition of DAPI allowed the cell cycle to be broadly quantified in terms of the proportion of cells in each phase of the cell cycle: G1, S and G2/M. A typical differentiation monitored using flow cytometry for the wildtype BOBSC cell line is shown in Figure 19a and b, and a schematic of the gene expression changes can be seen in Figure 19c. Permeabilised flow cytometry shows that the undifferentiated hiPSCs did not

express EOMES or SOX17, and a large proportion of the cells were in S phase of the cell cycle. During the differentiation process EOMES could be detected by 24 hours and was still present, although at a lower level, at 72 hours. SOX17 was expressed from 48 hours and most cells expressed the protein by 72 hours. The DAPI staining shows that during the differentiation process, there were a higher proportion of cells in G1 phase and there were fewer in S phase as expected from a less pluripotent cell type.

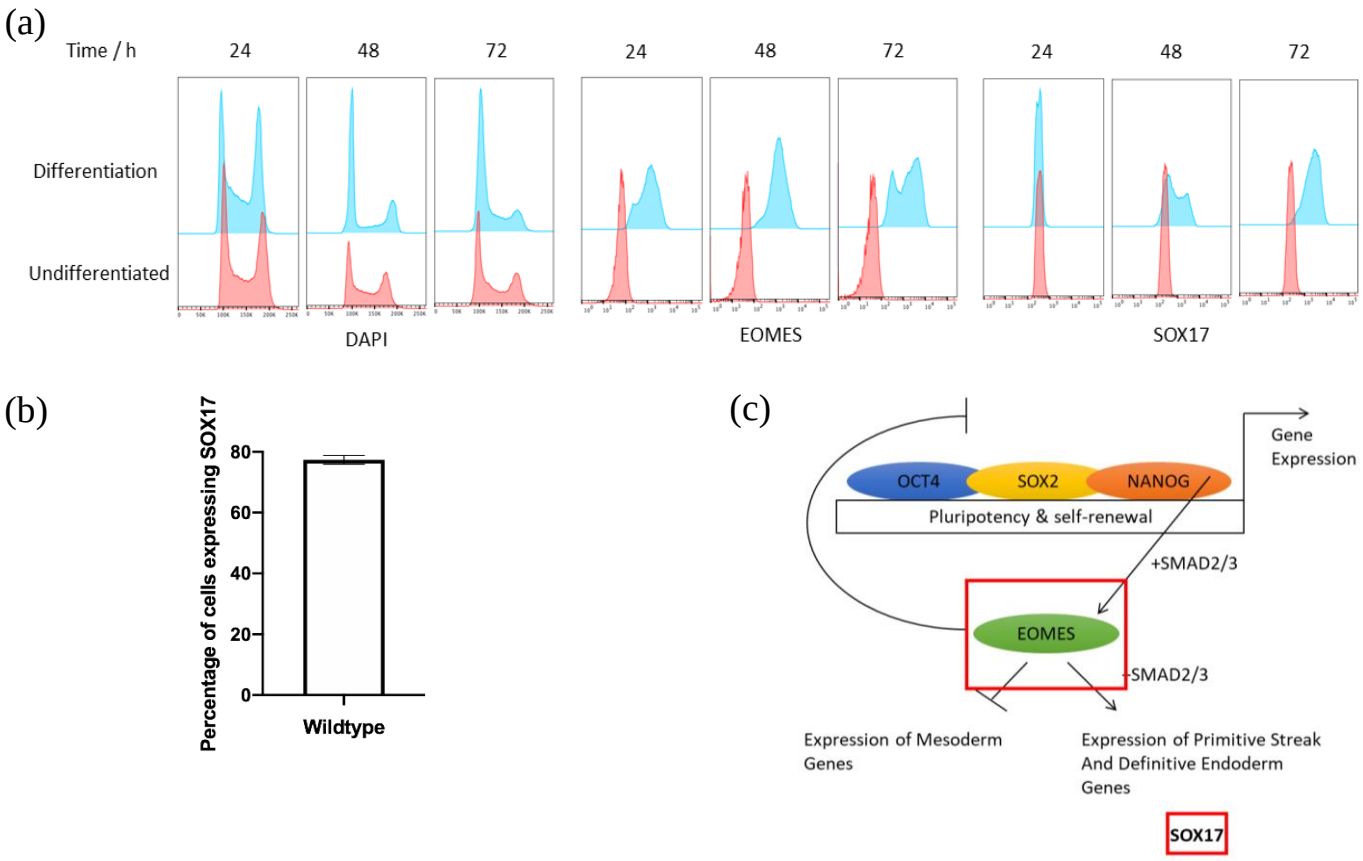


Figure 19. Definitive endoderm differentiation of hiSPCs (a) DAPI staining of DNA content of undifferentiated cells (blue) compared to differentiating cells (red) measured every 24 hours (left panel), EOMES staining (middle panel), SOX17 staining (right panel) (b) percentage of cells expressing SOX17 at 72 hours, n=20, mean±SEM are plotted (c) a schematic to show the gene expression changes during the differentiation process, proteins checked by permeabilised flow cytometry are marked with a red box.

The protocol required relatively few optimisations, but one major problem was selecting the most reliable inhibitors. The CHIR90021 inhibitor efficiency varied from batch to batch, altering the efficiency of the differentiation dramatically. After trialling several different manufacturers, the TOCRIS CHIR90021 was used as it gave the most consistent results, and this was added for all further experiments. The other main optimisation was the cell density and colony size of cells at the onset of differentiation, this has been shown to alter the efficiency of differentiation (Kempf et al., 2016). Splitting the cells 1:8 from full confluence 24 hours before

initiating differentiation consistently gave the highest proportion of SOX17 expressing cells and this was used throughout the experiments. As the confluency could alter the percentage of cells expressing SOX17 this was maintained as similarly as possible between cell lines and was kept the same between different drug-treatments. Once the differentiation was set up with high efficiency and reproducibility, the next step was to look at perturbing the differentiation. There were three ways in which to address this: G4-binding ligands as used previously, genetic knockouts of G4-processing enzymes and replication 'stress' including low doses of hydroxyurea (HU) and DNA damaging agents. I began by using G4 ligands as their addition had shown some suggestion of altering the differentiation in the mESC system.

3.2.3 G4-binding ligands

The first assessment of whether G4s may play a role in differentiation was using G4-binding ligands. The three G4 ligands previously used in the mouse cell experiments were tried on the hiPSCs as well as another G4 ligand, PhenDC3 (De Cian et al., 2007). The cells were grown in the presence of the ligands, PDS, NMM or PhenDC3 at 2 µM or PDC12 at 80 µM, in the undifferentiated state for a week to see whether they could be tested during the differentiation. NMM killed all cells within 48 hours even when the dose was taken down to the lowest active concentration in cells, 0.5 µM, therefore this drug was not used in the human system. The PDS and PhenDC3 could be tolerated by the cells for two weeks at 2 µM and these drugs were used for the experiments in this system. The PDC12, as before, exhibited a tendency to crystallise in solution and in order to induce epigenetic instability over five cell divisions, the length of the definitive endoderm differentiation process, the drug concentration would need to be around 80 µM. Since there was an adequate readout of the differentiation efficiency in this system, it was possible to test whether the volume of DMSO that would be added to the cells with PDC12 could affect the differentiation, regardless of the drug (Czysz et al., 2015; Pal et al., 2012). The cells were differentiated in the presence of DMSO to the equivalent of 80, 120 and 200 µM PDC12. The cells looked morphologically identical to the untreated cells but when the SOX17 percentage was analysed by flow cytometry it was clear that the higher the volume of DMSO added, the lower the proportion of SOX17 positive cells at the end of the differentiation process (Figure 20). While it was very interesting that high volumes of DMSO could alter the differentiation process so dramatically, it also meant that PDC12 could not be used on the human cells. Fortunately, each of the other drugs could be used at a much lower concentration whereby the DMSO was shown to have no effect (≤ 10 µM of a 10 mM solution). However, the control cells were usually vehicle-treated in these differentiation experiments. The results of this work are discussed in Results III.

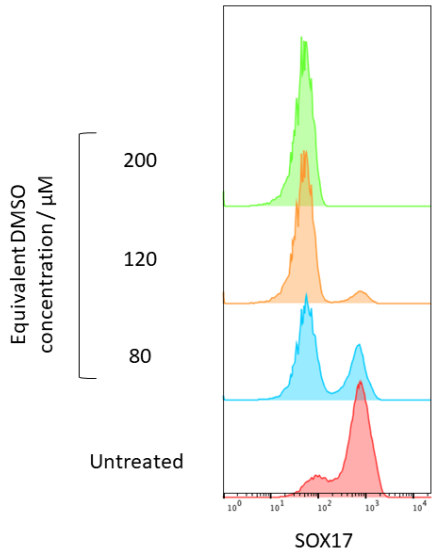


Figure 20. Analysis of the efficiency of definitive endoderm differentiation of hiSPCs in the presence of increasing volumes of DMSO with concentrations added equivalent to dilution from 10 mM stock (equivalent to the maximum solubility of PDC12): 0, 80, 120 and 200 µM. The histogram shows a decrease in SOX17 expression with increasing volumes of DMSO added.

3.2.4 G4-processing enzyme genetic knockouts

The Sanger Centre had generated many genetic knockouts in the BOBSC cell line for the COMSIG project, including *REV1*, *WRN* and *PRIMPOL*. This provided an opportunity to compare RNA sequencing data between these cell lines and wildtype cells treated with G4 ligands, as well as combine the ligands with the knockout cells lines to see if this exacerbates any phenotype. This data is discussed in Results IV.

3.2.5 Replication stress

Low doses of HU have previously been shown to induce epigenetic instability by means of nucleotide pool depletion, in a similar manner to G4 ligands and genetic knockouts of *REV1* and *PRIMPOL* in DT40 (Papadopoulou et al., 2015). The low doses of HU can therefore lead to uncoupling of the DNA replication fork and the processive helicase, and this can cause epigenetic instability. Theoretically, DNA damage, another form of replication stress, could cause epigenetic instability by a similar mechanism: a lesion is formed in the DNA which the leading strand polymerase cannot process through and therefore histone deposition and leading strand DNA synthesis are not coupled. How replication stress can influence maintenance of the ground state and the differentiation process will be addressed in Results II.

3.3 Neuroectoderm differentiation

Having set up the endoderm differentiation protocol, another differentiation pathway from hiPS cells was performed to understand whether similar results could be seen in a different lineage specification pathway. Neuroectoderm differentiation takes longer than endoderm

differentiation, around seven days, and can be continued to neuronal differentiation. It is a very robust protocol and the number of cell cycles gives sufficient time to allow a replication dependent phenomenon to become visible. Cells were set up for neuroectoderm differentiation in a similar manner to endoderm differentiation and the cell media was changed daily. For the first 48 hours cells were cultured in CDM-PVA, as in endoderm differentiation, supplemented with FGF2, CHIR99021, LDN193189 and SB431542. From day three to six cells were cultured in N2B27 medium containing SB431542. At day seven the SB431542 was removed from the N2B27 medium and cells were kept in this state until analysis. The LDN193189 was added to inhibit BMP type I receptors and SB431542 to inhibit TGFβ. Both are small molecule inhibitors of SMAD signalling and this is referred to as the LSB protocol (Menendez et al., 2011). Cells differentiated down this pathway exhibited much lower cell death when compared to the endoderm differentiation. The efficiency was measured using permeabilised flow cytometry analysing PAX6 and SOX1 expression.

A number of optimisations were required for this protocol: at day eight of differentiation the cells were very confluent, and the layer of adherent cells peeled off the plate so the differentiation could go no further. The cells were differentiated on gelatin instead of vitronectin to assess whether this could increase their attachment but there was further decreased attachment on gelatin. As before, the initial cell density was optimised to increase the efficiency of the process. The initial differentiation did not show any upregulation of PAX6 or SOX1 compared to the undifferentiated cells and therefore different batches of LDN193189 (Source BioScience) were trialled. Upon using this new inhibitor, and varying the initial cell density, the permeabilised flow cytometry showed an increase in both SOX1 and PAX6 expression compared to the undifferentiated state (Figure 21). While the shift in PAX6/SOX1 expression was not as distinct as the SOX17/EOMES expression in definitive endoderm cells verses undifferentiated state, there was a change. While optimising this protocol, the cells were differentiated to neuroectoderm in the presence of the G4 ligands PDS and PhenDC3 as discussed previously. Using 2 μM of each ligand showed both PDS and PhenDC3 caused the cells to start to die within three days and by the fourth day almost all the cells treated with the ligands were dead. Therefore, the remainder of the thesis is focused on definitive endoderm differentiation.

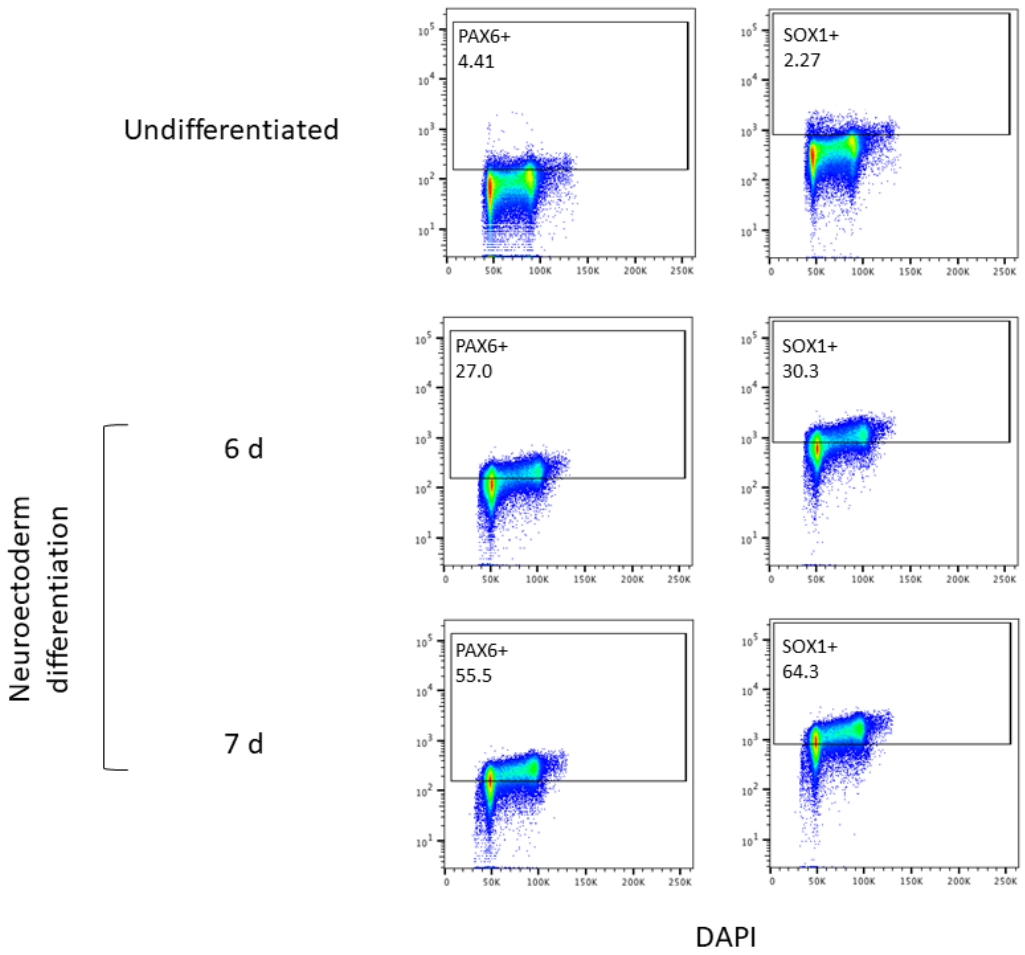


Figure 21. Neuroectoderm differentiation of hiPSCs stained for PAX6 (left panel) and SOX1 (right panel) and DAPI, in the undifferentiated state (top panel), after six days (middle panel) and after seven days (bottom panel). Gates are measured at the percentage of cells expressing PAX6 or SOX1.

Chapter 4. Results II: The DNA damage response in the undifferentiated state and during endoderm differentiation of hiPS cells

4.1 Changes to DNA damage markers in the pluripotent state

4.1.1 The unperturbed undifferentiated state

Early in the mES cell experiments, it was very clear that the cells could tolerate G4 ligands better in the ground state than during differentiation. This effect was more pronounced when comparing the hiPSC undifferentiated and definitive endoderm differentiating cell states, introduced in Results I and expanded on in Results III. Since the G4 ligand PDS has been shown to induce DNA damage (Rodriguez et al., 2012), the response to DNA damage in the undifferentiated and differentiating states are described in this chapter.

4.1.1.1 2D-cell cycle plots in hiPSCs

To give an insight into the cell cycle of these highly proliferative cells, undifferentiated, unperturbed replicating hiPSCs were cultured in the presence of the nucleotide analogue, EdU, for one hour and then fixed for flow cytometry analysis (Figure 22a). This revealed an unexpected 2D-cell cycle plot characterised by the presence of a late S phase population, possibly indicative of late damage repair, that also continued into the differentiation. This population of cells did not show an increase in γH2A.X (Figure 22b) or pRPA (Figure 22c) signal using flow cytometry. Therefore, it was not clear what this specific population was acting as and whether it was related to resolving DNA damage in late S phase in these highly proliferating cells.

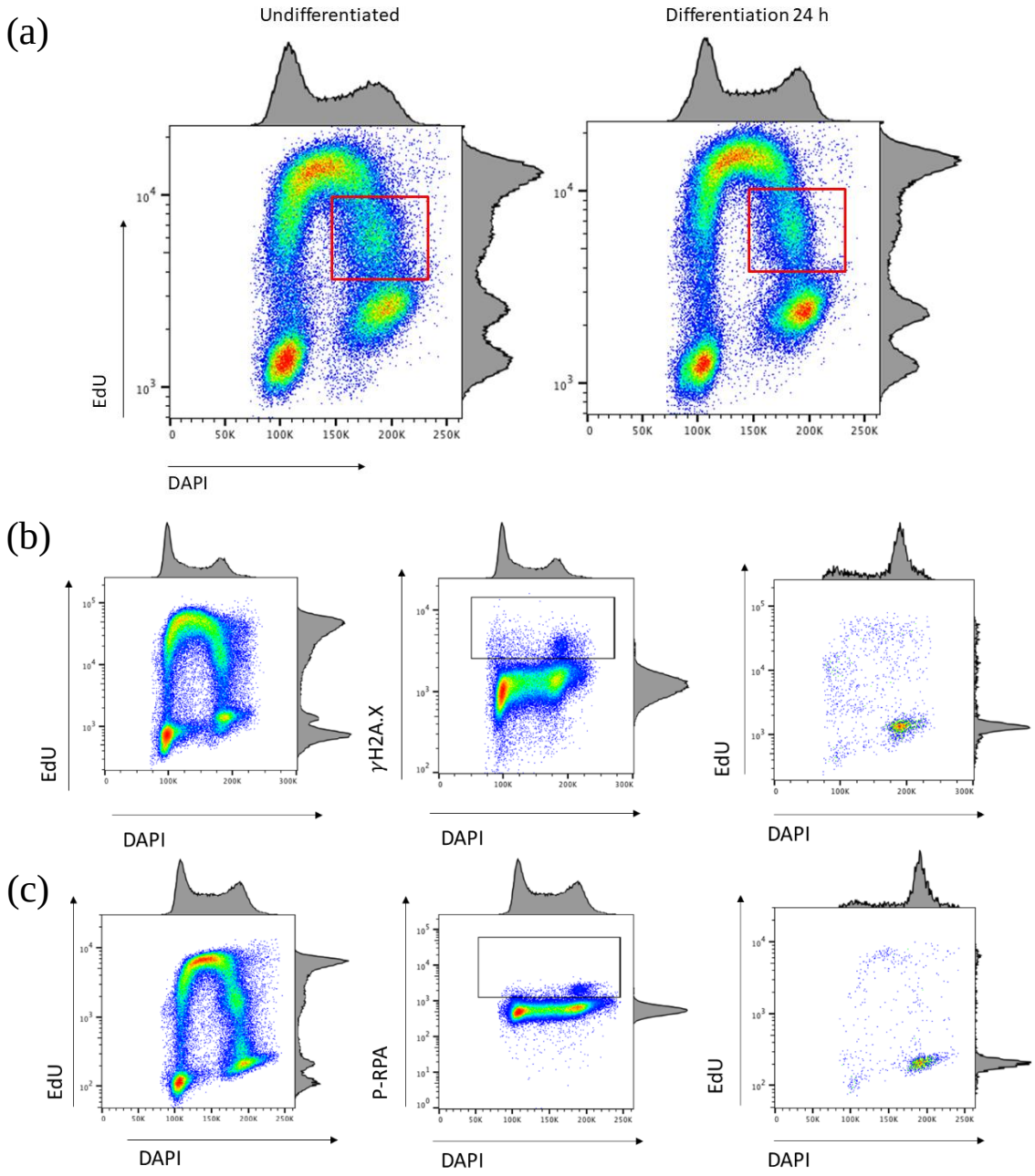


Figure 22. 2D-cell cycle analysis of wildtype BOBSC cells (a) EdU / DAPI plot of undifferentiated (left) and differentiating cells (right) to show the late S phase cell cycle bulge (shown in the red box on both plots) (b) gating of γH2A.X positive cells (positive cells shown in the middle plot with the black box) showing that these do not correlate with this bulge (c) gating of pRPA positive cells (positive cells shown in the middle plot with the black box) and showing there is no increase in this population. Adjunct histograms are shown.

4.1.2 Perturbing replication in the undifferentiated state

The aim of this thesis was to understand whether perturbing replication could impact differentiation. In order to address this question, it must be first understood how these perturbations impact the undifferentiated cell state.

4.1.2.1 Nucleotide depletion using hydroxyurea

It has been shown previously in the group that a low dose of hydroxyurea (HU) is able to induce epigenetic instability at the BU-1 locus, in a similar manner to G4 ligands and genetic knockouts of enzymes involved in processing G4s (Papadopoulou et al., 2015). This phenomenon occurs due to the hydroxyurea causing nucleotide pool depletion and replication stress (Alvino et al., 2007). Cells were grown in 50-150 µM HU without any observed changes in cell doubling time for five days; the cells could be maintained at these concentrations for over two weeks. Cells in the undifferentiated state were able to tolerate high doses of HU, again suggesting that the hiPSCs were capable of processing high levels of DNA damage. However, as very little change was seen in these cells the concentrations were increased to 300-500 µM where there was a slowing of growth (Figure 23). At the high dose of 500 µM, there was a large increase in the proportion of cells in S phase, as would be expected on treatment with HU, but very little cell death. Nucleotide pool depletion was used during definitive endoderm differentiation as discussed in Section 4.2.2.

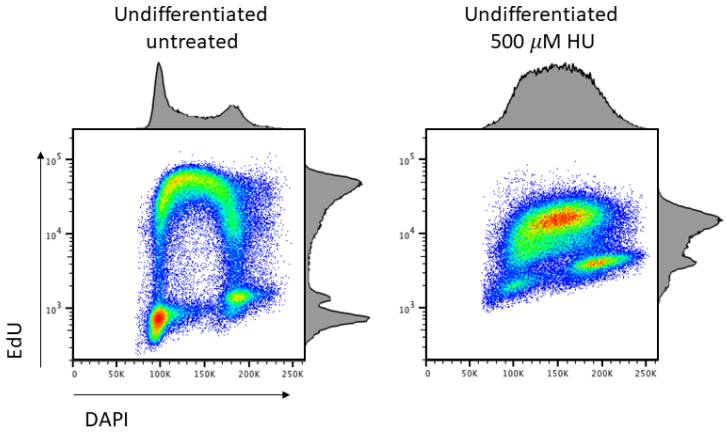


Figure 23. 2D-cell cycle plots of undifferentiated BOBSC cells cultured with 500 µM HU for 50 hours. The x-axis shows DNA content stained with DAPI and the y-axis shows incorporation of EdU. 50 hours was used as a comparison to the changes during differentiation, discussed later.

4.1.2.2 The effect of inducing DNA damage in the undifferentiated state

In order to study the way in which undifferentiated cells deal with DNA damage, cells were treated with different DNA damaging agents, which would induce different lesions or breaks on the DNA, in the undifferentiated state. Methyl methanesulfonate (MMS) is a DNA alkylating agent that was added to induce replication stress and DNA damage. When added to aqueous solution at 37ºC it has a half-life of 9.5 hours (Kilbey et al., 2012). It predominantly methylates DNA on N7-deoxyguanosine and N3-deoxyadenosine. In order to test the dose of MMS that could be tolerated without inducing excessive cell death or significant slowing of cell growth, the drug was added to undifferentiated cells at concentrations in the range of 0.25-10 ppm. Cell media was changed every 24 hours and new MMS was added, and after 72 hours the cells were checked with a microscope to analyse the rate of proliferation compared to untreated cells. At concentrations above 10 ppm there was a decrease in the rate of cycling visible by eye. Cells

treated with 2.5-5 ppm MMS showed very little difference in growth rate compared to untreated cells and 5 ppm was used for further experiments. MMS was added every 24 hours for three days and the cell cycle profile was analysed (Figure 24). There was very little cell death accompanied by smalls increases in the proportion of cells in G2/M phase of the cycle. These doses were used to perturb endoderm differentiation, discussed in Section 4.2.3.1.

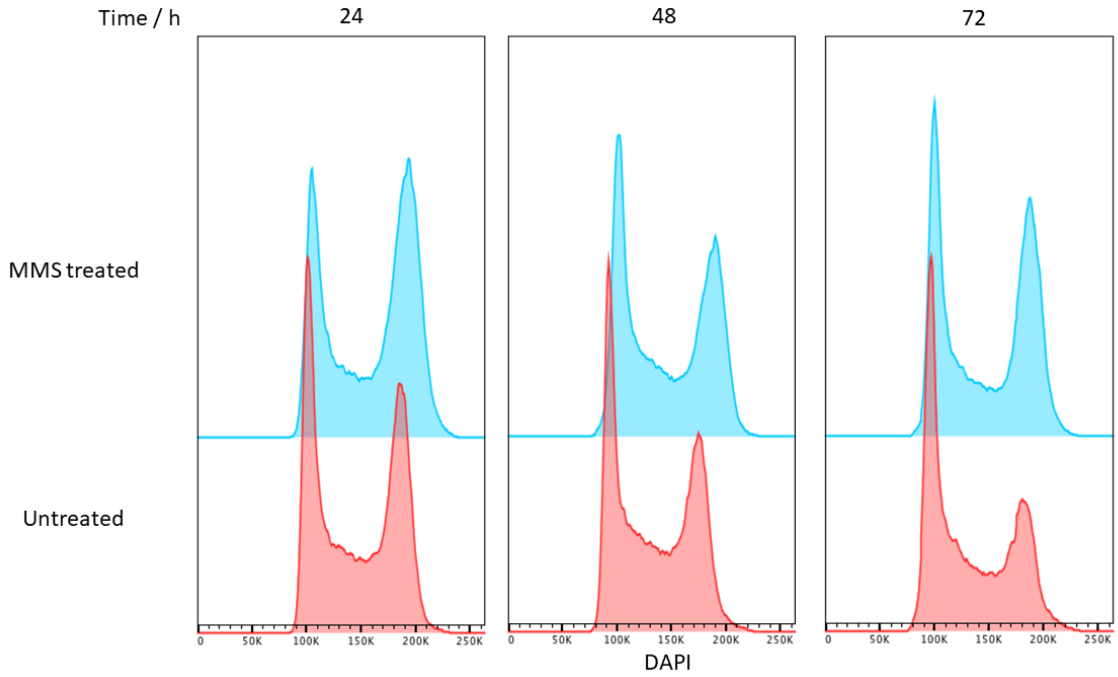


Figure 24. Flow cytometry plots to show DAPI staining of DNA content in undifferentiated cells treated with 5 ppm MMS (blue) at 0-, 24- and 48-hours and collected every 24 hours to check the changes in the cell cycle compared to untreated cells (red).

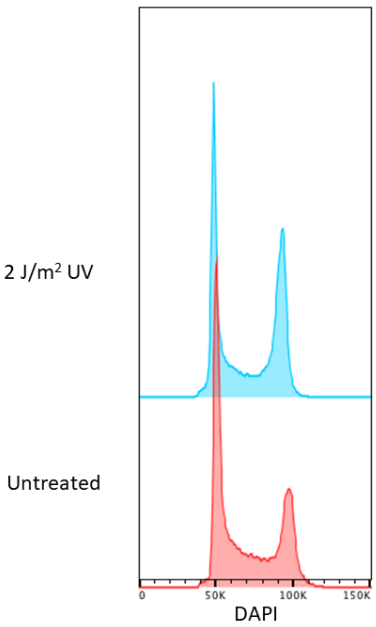


Figure 25. Flow cytometry plots to show the distribution of cells in the cell cycle after irradiation with 2 J/m^2 UV-C (blue) compared to untreated (red). Cells were treated at 24 hours and collected 48 hours later, this was chosen in order to compare to differentiated cells discussed further in Section 4.2.3.2.

Irradiation with UV-C was used to induce cytotoxic DNA lesions including cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs). A range of UV-C doses were trialled, and the highest tolerated dose was used further, as for MMS. Two joules per square metre of radiation was used as this dose caused very little cell death in the undifferentiated state but still caused an increase in the proportion of G2/M phase cells after 48 hours, typical of a replication impediment (Figure 25). The cells were treated at 24 hours, allowing comparison to the differentiated experiments; the choice of timepoints will be discussed later in this chapter with reference to differentiation (Section 4.2.5).

4.1.3 Changes to DNA damage markers after damage in the pluripotent state

To understand whether the BOBSC cells responded in a similar manner to somatic cells on treatment with DNA damage, I assessed whether the levels of serine-139 γH2A.X increased when low doses of HU or MMS were added. Protein was collected after 51 hours of culture and analysed by western blotting (Figure 26a). This timepoint of 51 hours was used as an important comparison to differentiating cells, discussed later in the chapter. Cells were also analysed using flow cytometry (Figure 26b). It was clear that both treatments caused an upregulation in the phosphorylation of H2A.X, however this was more pronounced in HU treated cells: this will be discussed in further detail in Section 4.2.

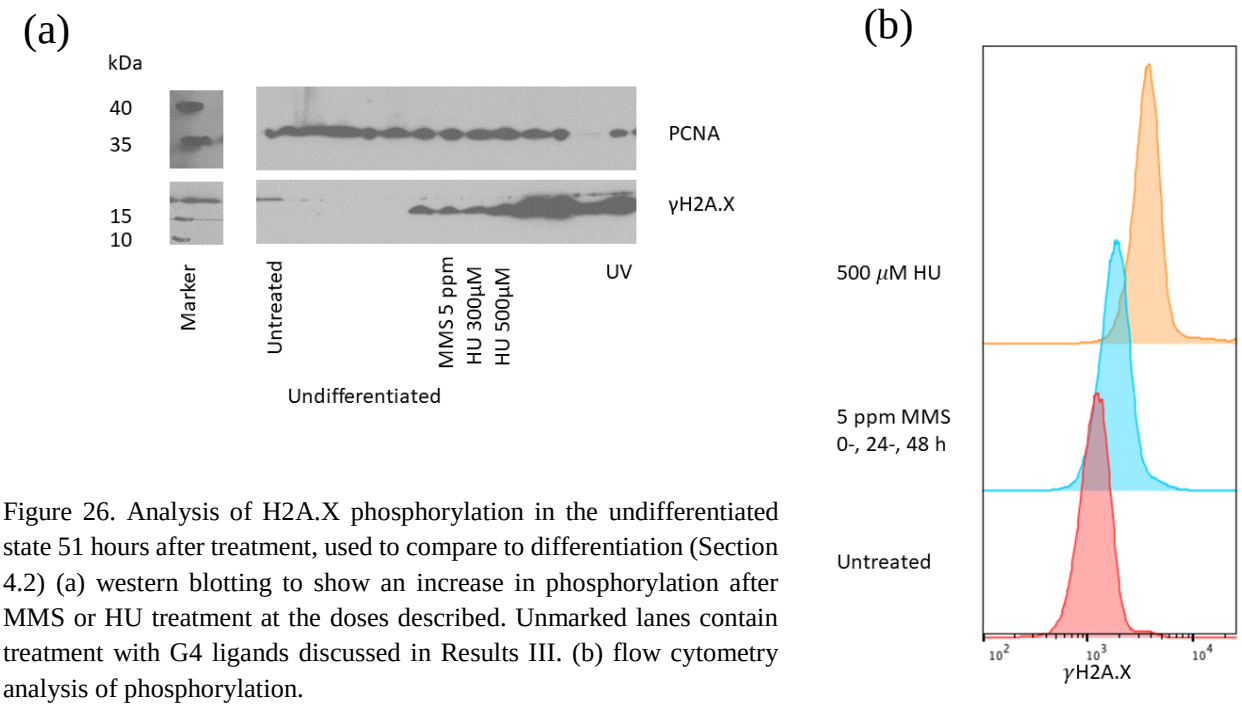


Figure 26. Analysis of H2A.X phosphorylation in the undifferentiated state 51 hours after treatment, used to compare to differentiation (Section 4.2) (a) western blotting to show an increase in phosphorylation after MMS or HU treatment at the doses described. Unmarked lanes contain treatment with G4 ligands discussed in Results III. (b) flow cytometry analysis of phosphorylation.

4.2 Changes to DNA damage markers during endoderm differentiation

4.2.1 Unperturbed differentiation

Before addressing how the DDR is triggered by DNA damage during differentiation, it was crucial to understand whether DDR markers change over the course of endoderm differentiation from undifferentiated cells. Therefore, levels of CHK1, CHK2 and RPA in undifferentiated, 24- and 48-hour differentiated wildtype cells were checked using western blotting (Figure 27a-c). Total CHK1, CHK2 and RPA proteins were expressed undifferentiated cells but the expression of CHK1 and CHK2 decreased during differentiation. This decrease was not seen with RPA. pCHK1, pCHK2 and pRPA signals were only seen in the positive control cells and were not changed during differentiation.

The tumour suppressor protein, p53, was also analysed by western blotting (Figure 27d). Total levels of p53 were high in undifferentiated cells, as for the other DNA damage markers, and its expression decreased during the differentiation process. It is known that p53 plays a role in the initiation of differentiation (Murray-Zmijewski et al., 2006). Testing the levels of p63 and p73 would have been useful as they are also known to play a role in differentiation (Wang et al., 2017). However, all antibodies tested did not give any signal and RNA sequencing discussed in Section 4.4 showed that the RNA of these transcripts were expressed at very low levels in the undifferentiated state and this continued throughout differentiation. These western blots suggested that a number of proteins involved in the DDR were expressed at high levels in the pluripotent state and these levels decreased throughout differentiation.

Phosphorylation of the histone variant H2A.X was also analysed as a marker of DNA damage using western blotting and flow cytometry. It was clear that the level of γH2A.X was inversely correlated with total CHK1 and p53 protein (Figure 27a, c, d). Undifferentiated cells had low levels of H2A.X phosphorylation, whereas at 48 hours into differentiation this γH2A.X signal increased to much higher levels compared to the undifferentiated state. This finding was entirely unexpected, as this marker is usually associated with DNA damage, and it was of interest to study what might be causing the increase at 48 hours into differentiation.

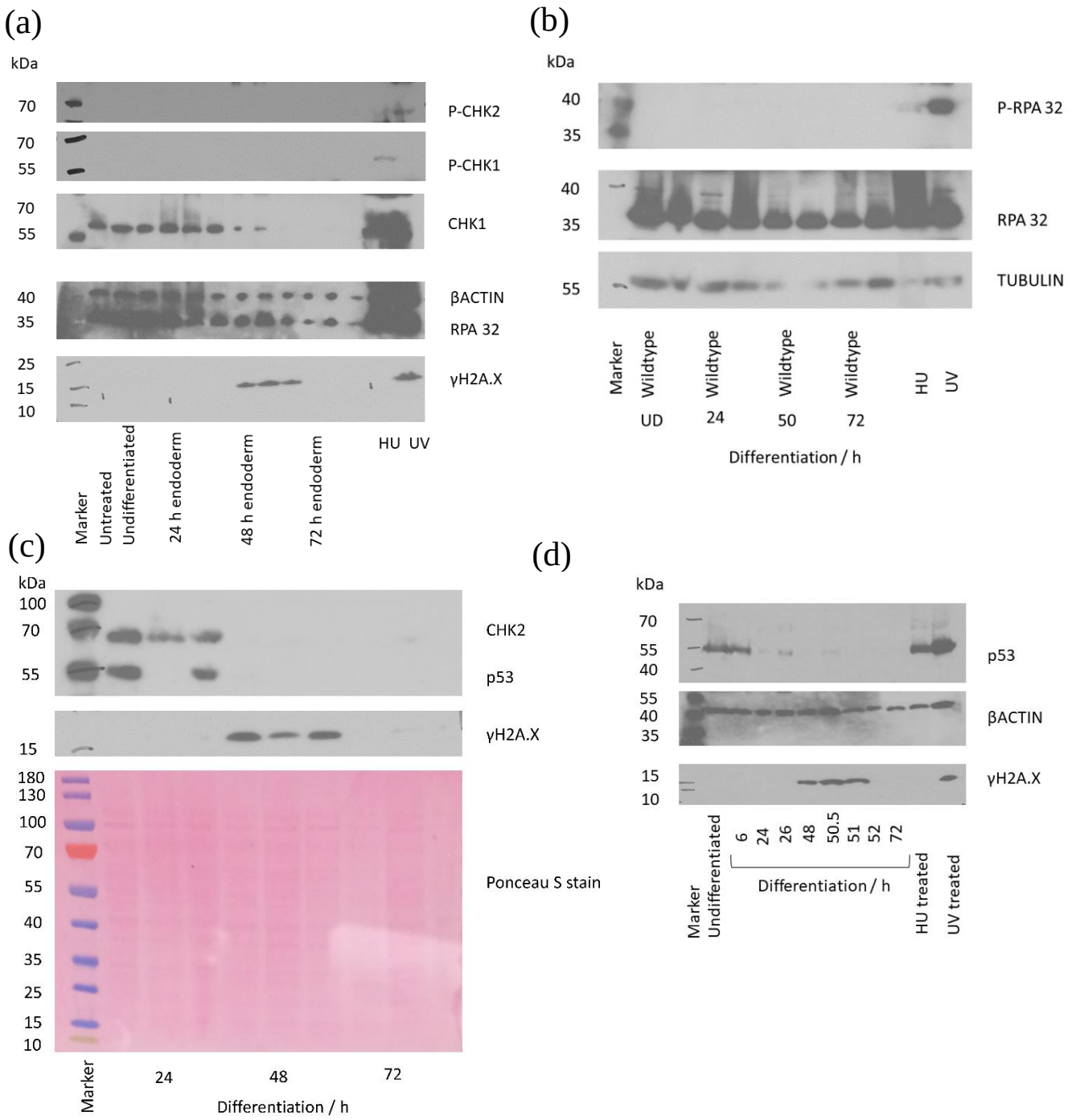


Figure 27. Western blotting analysis of proteins involved in the DDR during differentiation of BOBSC cells (a) CHK1 and CHK2 phosphorylation, total CHK1, RPA and γH2A.X in the undifferentiated state and every 24 hours of differentiation (b) RPA phosphorylated and unphosphorylated state in the undifferentiated state and every 24 hours of differentiation (c) total CHK2 and p53, and γH2A.X every 24 hours of differentiation (d) total p53 and γH2A.X during differentiation. Unmarked lanes contain conditions not discussed in this chapter. UD is undifferentiated. UV and HU treatment were used as positive controls.

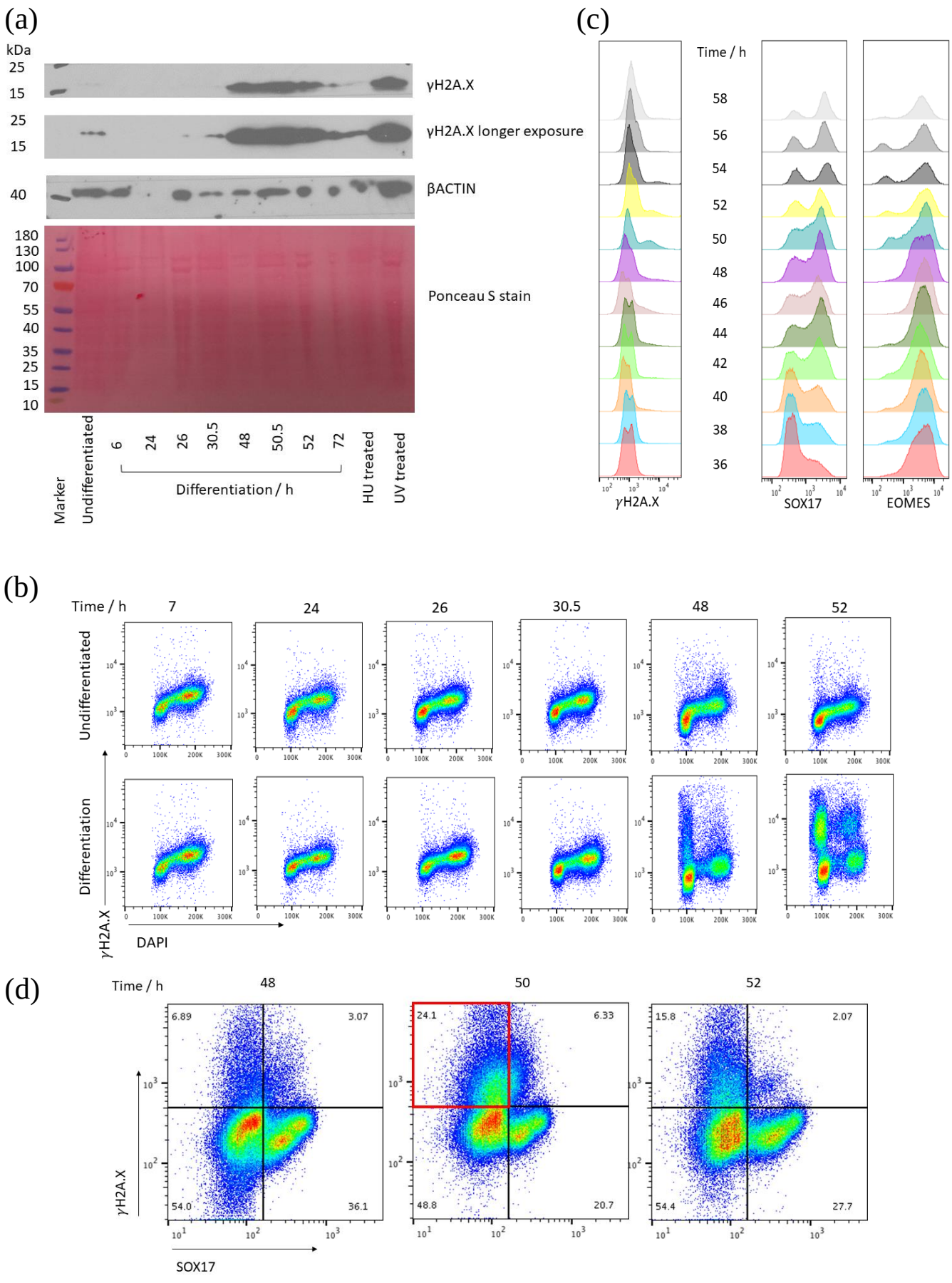
4.2.1.1 Phosphorylation of H2A.X during unperturbed differentiation

To gain further insight into when the high level of γH2A.X was appearing during the differentiation, protein was collected across more intervals and western blotting was performed (Figure 28a). This shows that the marker was present from at least 48 to 52 hours into the differentiation. To look in more detail at an individual cell level, γH2A.X flow cytometry during differentiation was performed. Cells were collected from 7 hours to 52 hours throughout the differentiation process, as well as in the undifferentiated state, and analysed for γH2A.X (Figure 28b). This experiment showed a clear increase in γH2A.X at both 48 and 52 hours. Interestingly the cells exhibiting this marker were mainly in G1 phase with some in G2/M. This was again unexpected because γH2A.X is normally seen in S phase cells. This suggests that what was seen may not have been the typical DNA damage response pathway.

In order to ascertain at which stage of the differentiation H2A.X was phosphorylated, cells were collected every two hours from 36 to 58 hours and analysed using flow cytometry to check for both γH2A.X and SOX17 individually (Figure 28c). This showed that a small proportion of cells started upregulating γH2A.X from 38 hours and this continued until 58 hours: the highest levels were monitored from 46 to 54 hours, as previously seen. This increase in γH2A.X did not correlate with the onset of SOX17 expression, or a decrease in EOMES expression and was therefore unlikely to be associated with the expression of these genes. However, this time window did seem to correlate with the period in which the cells were believed to transition from being epithelial to mesenchymal.

To confirm that there was no dependency of SOX17 upregulation on H2A.X phosphorylation, or vice versa, flow cytometry was performed to check individual cells for both SOX17 and γH2A.X expression (Figure 28d). The majority of cells expressing SOX17 did not have a positive γH2A.X signal, and vice versa, suggesting that upregulation of both may have been mutually exclusive. These cells may need to have H2A.X phosphorylated and then dephosphorylated before expressing SOX17, such that each cell goes through a transition of having H2A.X phosphorylated before specifying differentiation but confirming this would require live cell imaging and be almost impossible to monitor.

Figure 28. (See next page). Analysis of the phosphorylation of H2A.X during definitive endoderm differentiation (a) western blotting analysis (b) flow cytometry analysis (c) flow cytometry analysis with a comparison to EOMES and SOX17 expression (d) comparison of the level of γH2A.X and SOX17 on a per cell basis, the red box shows cells negative for SOX17 expression but positive for H2A.X phosphorylation.


(a)
kDa
25
15
25
15
40
180
130
100
70
55
40
35
25
15
10
γH2A.X
γH2A.X longer exposure
βACTIN
Ponceau S stain
Marker
Undifferentiated
6
24
26
30.5
48
50.5
52
72
Differentiation / h
HU treated
UV treated
(c)
Time / h
58
56
54
52
50
48
46
44
42
40
38
36
γH2A.X
SOX17
EOMES
(b)
Time / h
7
24
26
30.5
48
52
Undifferentiated
Differentiation
γH2A.X
DAPI
(d)
Time / h
48
50
52
6.89
3.07
54.0
36.1
24.1
6.33
48.8
20.7
15.8
2.07
54.4
27.7
γH2A.X
SOX17

To verify whether the total levels of H2A.X protein correlated with the level of phosphorylation, H2A.X was monitored using flow cytometry (Figure 29a and b) and western blotting (Figure 29c). The flow cytometry data, while showing a greater spread in the level of H2A.X, did not show an increase in total H2A.X protein comparable to the phosphorylated variant, although the western blot did give a signal between 48 and 51 hours. This upregulation was not as great as the phospho-H2A.X spike, but since it was at an identical time it is not possible to rule out the possibility that the level of total protein was responsible for this difference. However, the antibody may be non-specific and could have a preference for phosphorylated H2A.X, it is expected to bind to the C-terminal region of the peptide, as the migration of the phosphorylated and unphosphorylated proteins appear very similar from the blots.

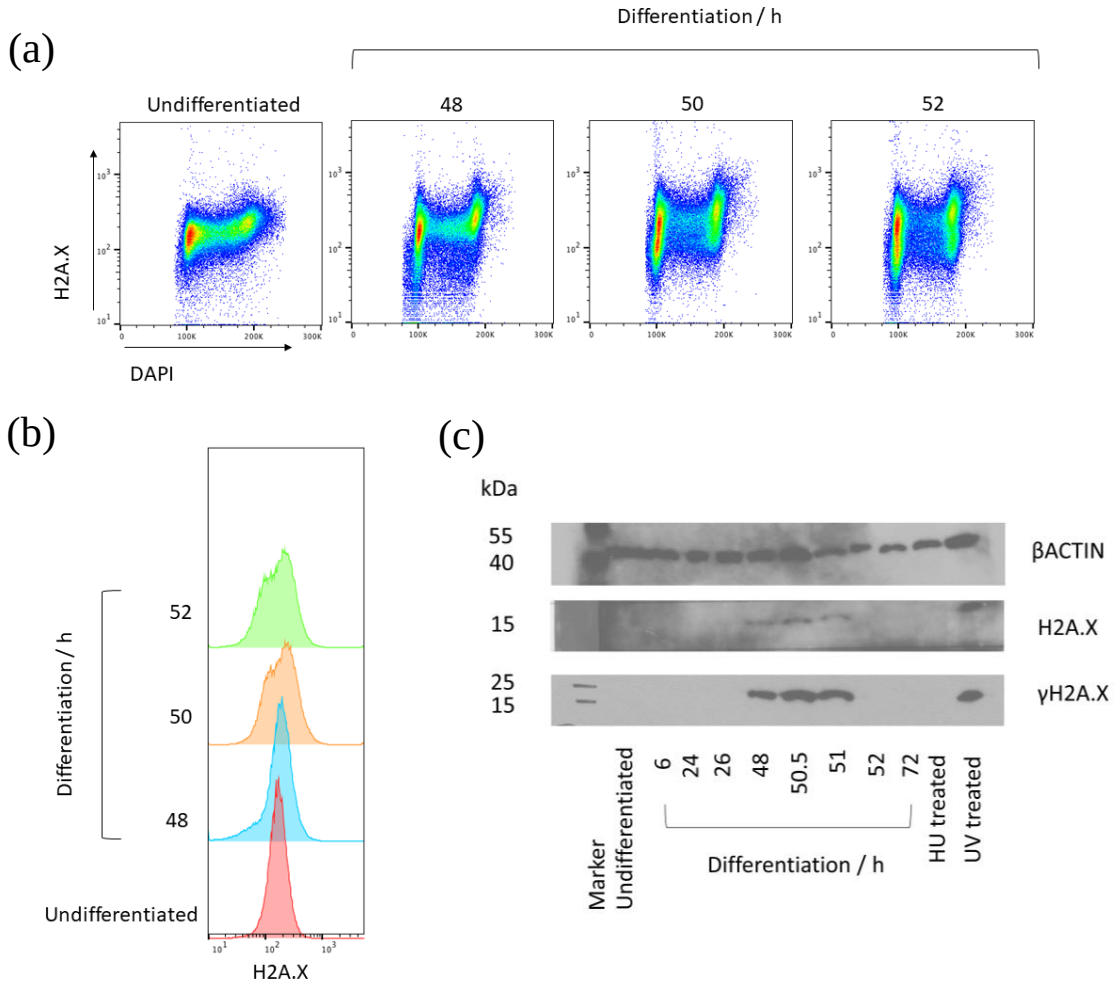


Figure 29. Analysis of total unphosphorylated H2A.X at the time of the spike of phosphorylation during differentiation (a) flow cytometry analysis of H2A.X with DAPI staining (b) as in (a) but showing total H2A.X only (c) western blotting analysis of H2A.X phosphorylated and unphosphorylated.

Previous work on the epithelial mesenchymal transition (EMT) in cancer has proposed a mechanism for H2A.X phosphorylation mediated by ATM and the SRY-box protein HMG2A (Singh et al., 2015). As H2A.X is thought to be able to be phosphorylated by DNA-PK, ATR

and ATM (Burma et al., 2001), inhibitors of ATM and ATR kinases, KU-60019 (Selleckchem) and AZD6738 (Selleckchem) respectively, were used in an attempt to understand which kinase was phosphorylating H2A.X during differentiation. Previous publications had used 10 μM of each inhibitor, and I tested this concentration in the undifferentiated state but this killed the cells in under 48 hours. Therefore, this approach was not able to answer the question, possibly because ATM and ATR are required in the pluripotent state. Rather, the ATM inhibitor was used at 2 μM and the ATR inhibitor at 1 μM as the undifferentiated cells could tolerate these concentrations for one week. Cells were differentiated or kept in the undifferentiated state in the presence of either one or both of the drugs. However, the differentiating cells reacted very poorly to the drugs and there was an increased level of cell death with each treatment. The cells were differentiated for 48 hours without inhibitors and 24 hours with inhibitors to minimise the time grown in the presence of the inhibitor, either 0-24, 24-48 or 48-72 hours, and assessed for SOX17 expression (Figure 30a). Adding the ATM inhibitor in the first 24 hours had no effect on the population of SOX17 expressing cells but all other conditions decreased the differentiation efficiency and therefore these pathways may be required during differentiation. The γH2A.X spike seen did not appear until much later than 24 hours, which could explain these observations.

I next investigated whether the ATM or ATR inhibitors were able to alter the proportion of γH2A.X positive cells. To minimise the toxic effect caused by the inhibitors, cells were grown for four or 20 hours on either one or both of the inhibitors at a higher concentration of 5 μM prior to collection. Cells were differentiated to 52.5 hours and the level of γH2A.X was analysed using permeabilised flow cytometry (Figure 30b). However, this was not able to answer the question as most inhibitor treated cells had an increased percentage of γH2A.X positive cells. The only condition where there was a decrease in the phosphorylation was using the combination of inhibitors for four hours prior to collection but a small spike was still observed. A different ATM inhibitor KU-55933, that acts less specifically than KU-60019, and has been previously used when looking at inhibiting the EMT associated γH2A.X was used (Singh et al., 2015). The experiment was performed as before, with the drugs added four hours before collection, and cells were collected at 51.5 hours for permeabilised flow cytometry analysis (Figure 30c). However, there was an increase in the phosphorylation of H2A.X during differentiation compared to the other ATM inhibitor.

The use of the inhibitors of ATM and ATR kinase activity did not give any clear results, so an *ATM* knockout cell line generated by the Sanger Centre COMSIG project was used: *ATM*$^{-/-}$.

These cells differentiated with a similar efficiency to the wildtype cells as shown using flow cytometry (Figure 31a and b). This was expected because both *ATM*$^{-/-}$ human and mice are viable although display severe phenotypes (Barlow et al., 1996). The phosphorylation of H2A.X at the EMT also occurred in these cells (Figure 31c). This suggested that if ATM does phosphorylate H2A.X in this context, it is redundant in this pathway. Combining *ATM*$^{-/-}$ with the ATR inhibitor, as used above, the γH2A.X spike was analysed (Figure 31d). However, there was no suggestion that this decreased the level of phosphorylation, in fact there was an increase in the base level and the spike of phosphorylation in G2/M. The *ATM* knockout could not be clearly verified by western blotting (Figure 32), but the cells showed an increased sensitivity to bleomycin treatment (data not shown). RNA sequencing was performed but showed very few gene expression differences compared to wildtype cells, suggesting that this may not have been a true knockout cell line. In order to fully understand the mechanism by which H2A.X is being phosphorylated, a double knockout cell line of *ATR* and *ATM* could be used to see if this prevents phosphorylation. A triple knockout with *PRKDC* could be used if this approach yielded no results. Caffeine acts as a non-specific inhibitor of both ATR and ATM phosphorylation, treatment with this drug should prevent phosphorylation if these kinases are responsible.

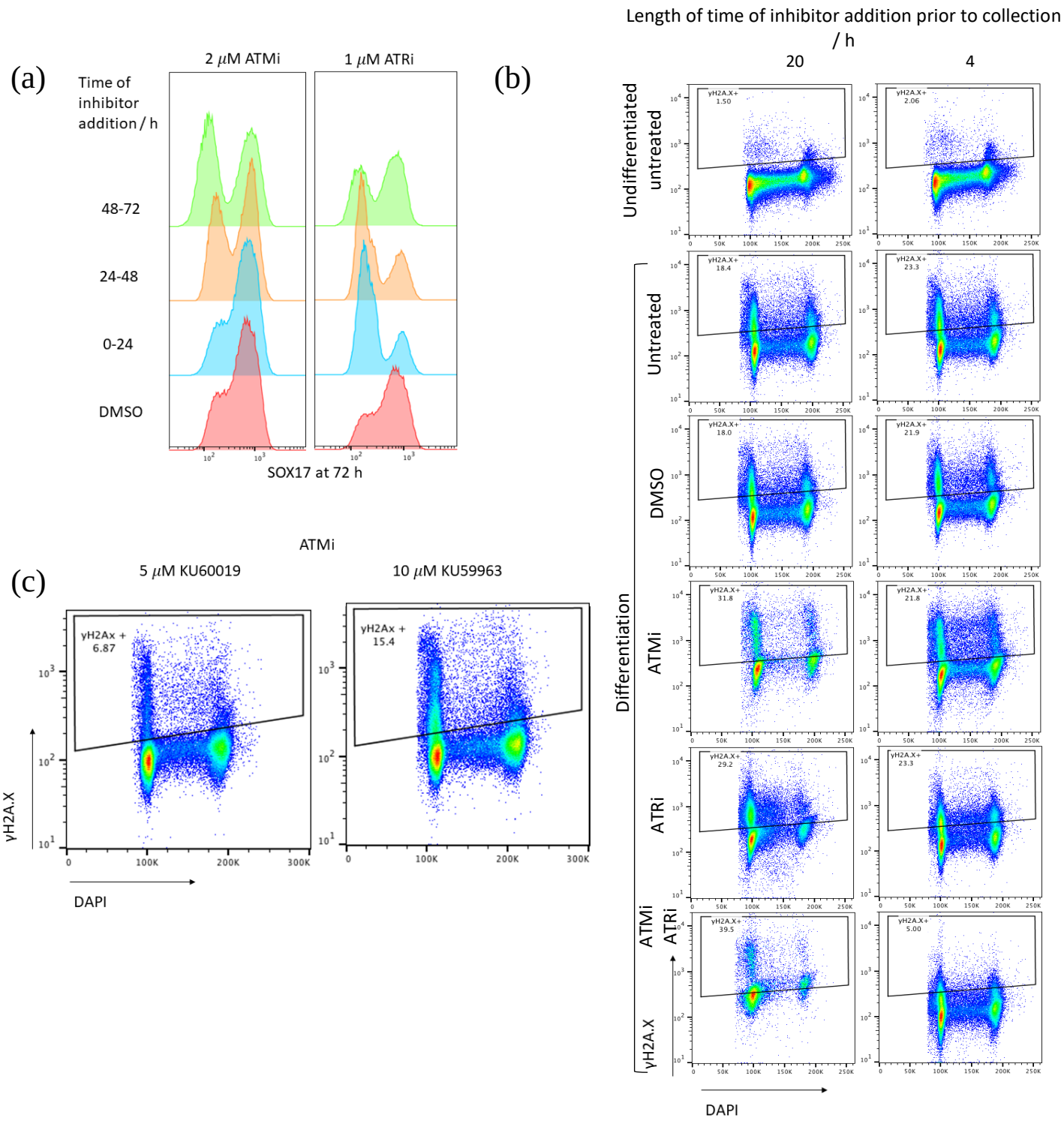


Figure 30. Flow cytometry analysis of ATM and ATR inhibitors (a) analysis of the cells expressing SOX17 at 72 hours after treatment with 1 µM ATR inhibitor or 2 µM ATM inhibitor for 24 hours during differentiation (b) 5µM of ATM, ATR or both inhibitors were added either 20 or 4 hours before collection at 52.5 hours to monitor the spike of phosphorylation (c) the phosphorylation of H2A.X at 51.5 hours after four hours of treatment with different ATM inhibitors. The baseline level of H2A.X phosphorylation changed with treatment, therefore the gating is relative to the untreated and not the baseline in each sample. This increase is likely to be due to increased cell death and damage.

.

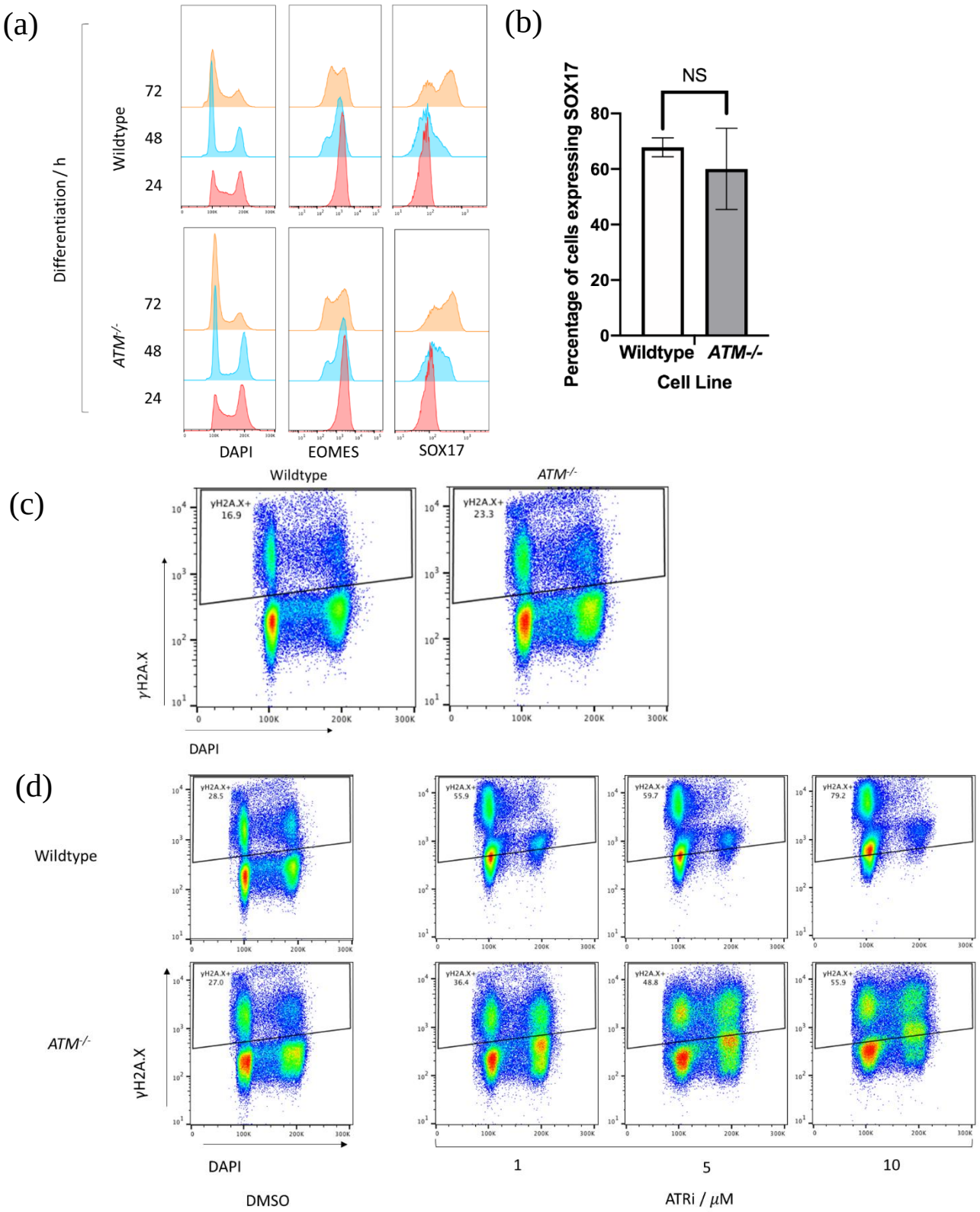


Figure 31. Flow cytometry analysis of differentiation of the *ATM*$^{-/-}$ cell line (a) analysis of the efficiency of differentiation every 24 hours monitoring DNA content, EOMES and SOX17 (b) replicate experiments n=2, mean±SEM are plotted, difference is not significant using an unpaired t test (c) the phosphorylation of H2A.X at 50 hours of differentiation compared to the wildtype cell line (d) use of the ATR inhibitor (Figure 30) in the *ATM*$^{-/-}$ cell line to monitor the level of H2A.X phosphorylation. As before, the positive gating is relative to the wildtype untreated and therefore the spike of phosphorylation above the baseline is not measured here. This is due to the inhibitors inducing damage in the cells, causing an increase in the baseline level of H2A.X phosphorylation. This spike of phosphorylation can be seen in all conditions.

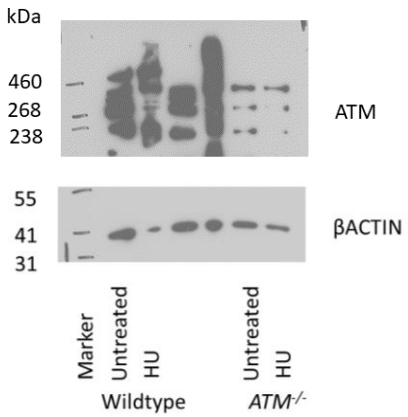


Figure 32. Analysis of the knockout of *ATM* in the *ATM*$^{-/-}$ cell line by western blotting, ATM should be seen at 350 kDa, unmarked lanes contain the *TP53*$^{-/-}$ cell line not discussed here, HU represents 2 mM hydroxyurea treatment for six hours prior to collection as a DDR positive control.

During the canonical DDR, H2A.X is phosphorylated during S phase of the cell cycle but in this data, the upregulation occurs in G1 and G2/M. To find out whether this upregulation was initiated in G1 or G2/M, the cells were synchronised in nocodazole in the undifferentiated state before being released into differentiation medium and initiating differentiation. The cells were collected at the point of the EMT spike (Figure 33). Nocodazole was used because it has been shown to affect stem cell differentiation efficiency the least of any means of cell synchronisation (Yiangou et al., 2018), however this synchronisation is only complete for one cell cycle. Nonetheless, the nocodazole synchronisation caused a large amount of cell death and the base level of H2A.X phosphorylation was also higher, particularly soon after release. The cells did not all release at identical times so the results were not definitive. The synchronised cells lagged in differentiation, probably due to their slow release, and the γH2A.X spike was not as great compared to the unsynchronised cells. However, the initial upregulation seen at 48 hours appeared to be in G1, suggesting that γH2A.X was upregulated in G1 phase of the cell cycle.

γH2A.X ChIPseq was performed on cells differentiated for 50 hours and cells maintained in the undifferentiated state. This was to understand whether this mark was associated with particular regions of the genome or a subset of genes, as it has been suggested to alter chromatin structure and transcription factor binding patterns at the EMT in cancer (Singh et al., 2015; Weyemi et al., 2016). The cells were collected for ChIP, and also for flow cytometry to check that the cells were differentiating (Figure 34). This figure shows that the cells were differentiating well and could be used for ChIPseq. The cells collected for ChIP were crosslinked, sonicated and the chromatin was spiked with HU treated DT40 chromatin at a 1:10 dilution. The immunoprecipitation was performed using an H3 antibody and a γH2A.X antibody. The DNA

extracted was used to make DNA libraries and which were sent to the CRUK sequencing facility. This data is being analysed at the moment and therefore is not in the thesis at this stage.

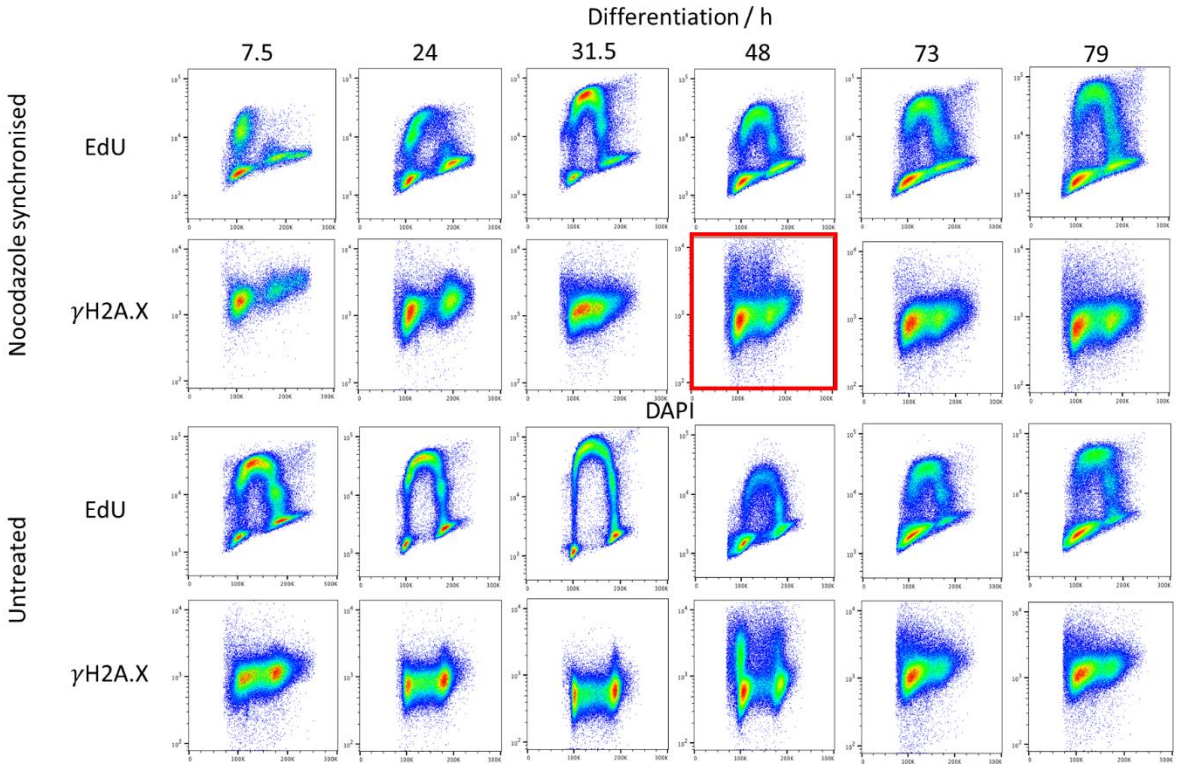


Figure 33. In an attempt to understand whether the phosphorylation of H2A.X was occurring during the cell cycle, cells were synchronised with 200 ng/mL Nocodozole for 16 hours in the undifferentiated state and when the Nocodazole was washed out the medium was replaced with the differentiation medium. EdU was added one hour prior to collection. The red box indicates synchronised cells at the point where H2A.X phosphorylation can first be seen in G1.

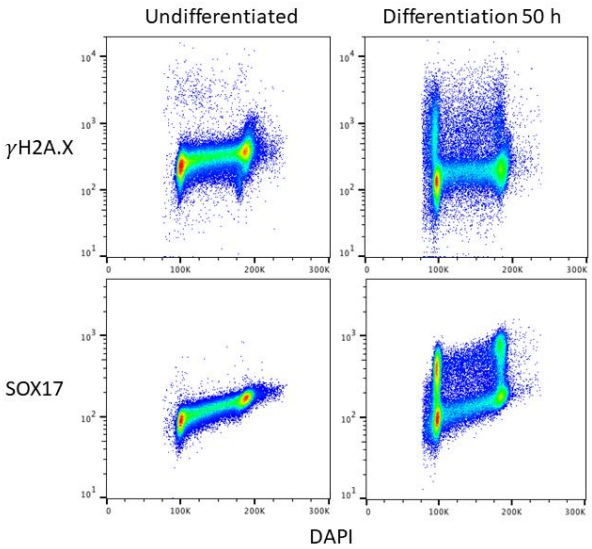


Figure 34. Flow cytometry analysis of the cells collected for ChIP analysis. Undifferentiated and differentiating cells were collected at 50 hours and analysed for SOX17 and phosphorylation of H2A.X.

4.2.1.2 The role of p53 in definitive endoderm differentiation

p53 is known to play a role in mesendoderm differentiation (Wang et al., 2017) and also to be downregulated at the EMT (Chang et al., 2011). It is also phosphorylated and activated for its role in the DNA damage response by ATM which was hypothesised to phosphorylate H2A.X at the EMT (Singh et al., 2015). Initial western blots, Section 4.2.1 (Figure 27c and d), suggested that total levels of p53 were high in the undifferentiated state and at the onset of differentiation. The levels then decreased throughout definitive endoderm differentiation. Phosphorylation of p53 at serine-15 was not detected unless cells were treated with MMS in either the undifferentiated or differentiating samples (Figure 35a and b). Preliminary flow cytometry data showed that cells which showed higher levels of γH2A.X at the spike of phosphorylation at 48 hours tended to express p53 to a lower level of SOX17 (Figure 35c). This data suggests that the level of p53 protein may be downregulated before H2A.X is phosphorylated at the EMT.

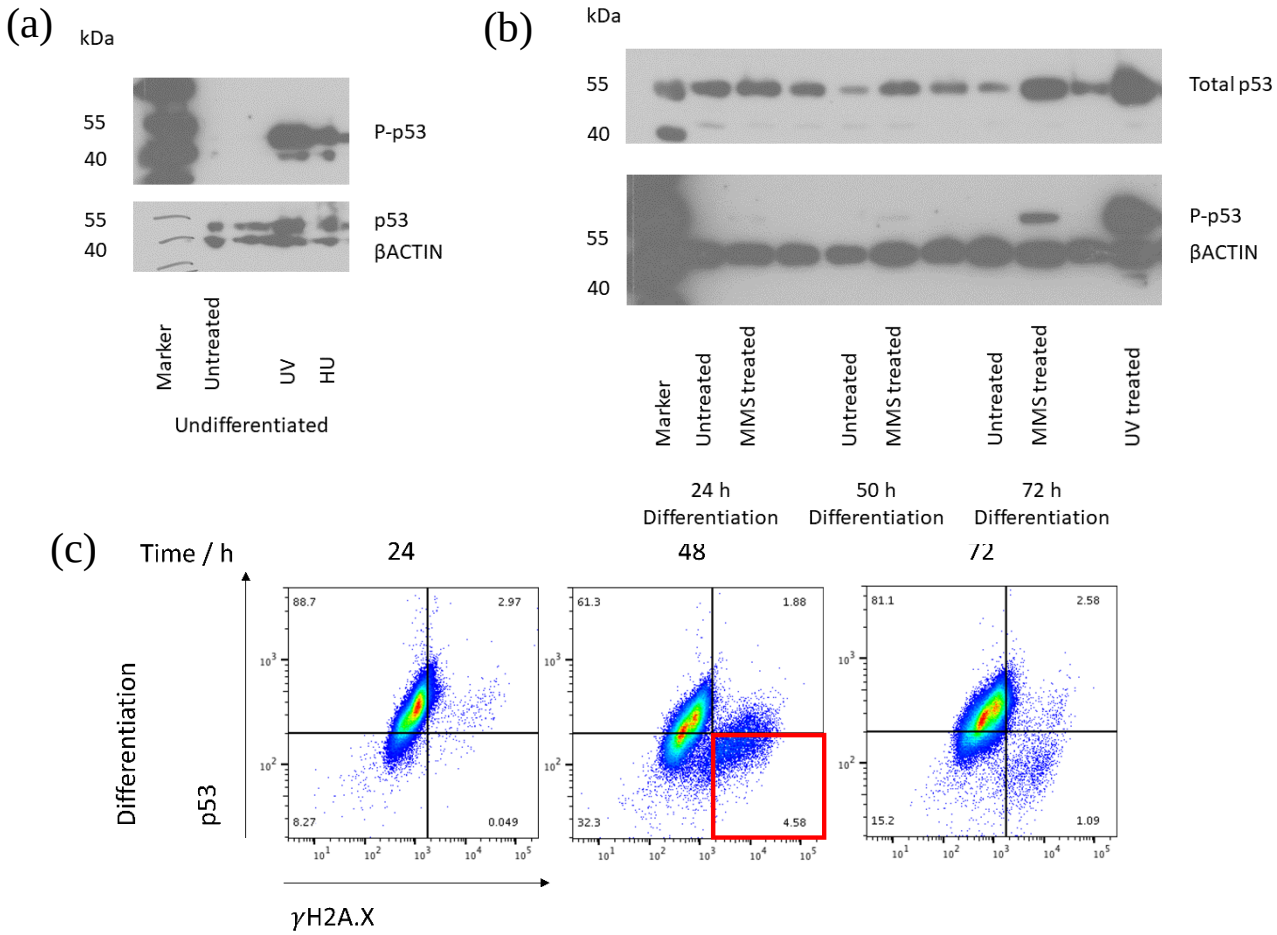


Figure 35. Analysis of p53 protein during differentiation (a) western blotting analysis to show phosphorylated p53 compared to total p53 in undifferentiated BOBSC cells (b) western blotting analysis of the levels of p53 and phosphorylated p53 during differentiation and after MMS treatment, unmarked lanes are G4 ligands not discussed here (c) the correlation of total p53 and γH2A.X during differentiation, the red box at 48 hours marks H2A.X phosphorylation and low levels of p53 at the spike of phosphorylation.

Because the levels of p53 decreased during differentiation, I attempted to stabilise p53 using the MDM2 inhibitor Nutlin-3a to probe whether this could inhibit the differentiation process. However, two brands of Nutlin-3a were trialled (Selleckchem and Cayman Chemicals) and both caused cell death within two days at the suggested effective concentration, so this approach was taken no further.

To check whether the total level of p53 protein could still be upregulated in response to high levels of DNA damage during the differentiation process, cells were treated with high doses of HU (2 mM) and collected for flow cytometry (Figure 36a). The level of p53 decreased during differentiation, especially at 54 hours, but increased on treatment with HU regardless of the time of treatment.

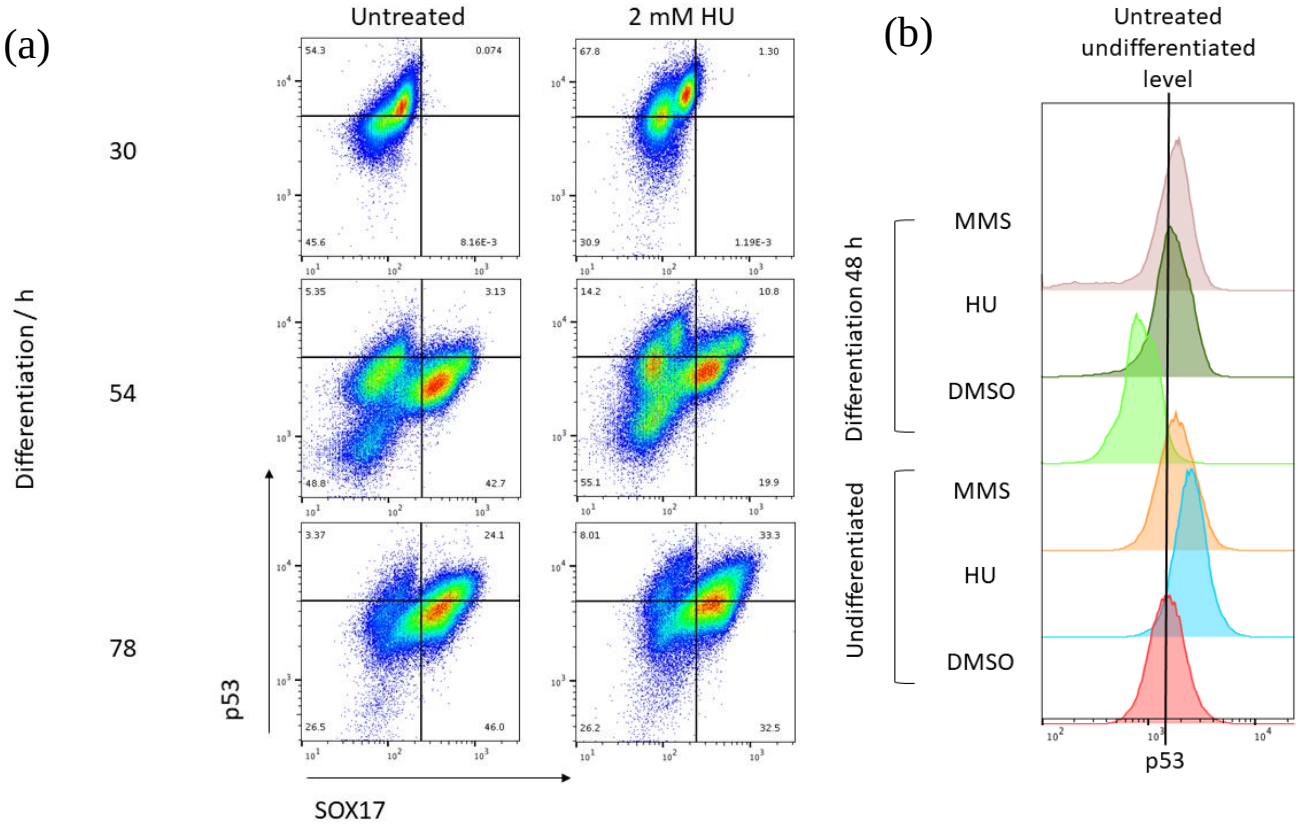


Figure 36. Flow cytometry analysis of the level of total p53 protein (a) in differentiating cells after treatment with 2 mM HU at 24, 48 and 72 hours. Cells were collected six hours after treatment (b) in undifferentiated and differentiating cells at 48 hours. Cells were treated with 5 ppm MMS or 300 µM HU at 24 hours and collected at 48 hours. The undifferentiated untreated level is marked.

It was clear that p53 levels were decreasing during differentiation and that this may coincide with the γH2A.X spike. Addition of high doses of HU during differentiation could also increase the level of p53. In order to understand how these levels corresponded to that in the undifferentiated state, permeabilised flow cytometry was performed to analyse the p53 signal, both with and without the addition of DNA damaging agents (Figure 36b). This shows that the level of total p53 protein was higher in the undifferentiated state than the differentiated state, but the level could be increased in both undifferentiated and differentiated cells after treatment with DNA damaging agents. When HU and MMS were added at 24 hours and the cells were collected at 48 hours, the p53 level was increased to a level equivalent to that of the undifferentiated state. This is discussed further in relation to differentiation perturbation in Section 4.2.4.3.

As *TP53* knockout mice are viable (Donehower et al., 1992), it was interesting to understand the role this protein plays in definitive endoderm differentiation and how it controls the DNA damage response throughout. p53 cells generated by the Sanger Centre COMSIG project were used. The knockout cells were verified using western blotting (Figure 37), which shows that the p53 antibody was binding a lower molecular weight protein, likely to be a truncated form of p53. This antibody binds to the N-terminal region, amino acids 20-25, and since the knockout is in exon six this is likely to be causing the difference. Therefore, while this is a knockout, it is clearly not of the whole protein: some of the p53 capabilities may still be functional. The targeting method states that the homologously recombined allele left only the transactivation domain in the N-terminus functional, whereas the CRISPR/Cas9 knockout allele acted in the DNA binding domain. This suggested that the transactivation, and therefore MDM2 binding domain, would certainly be intact.

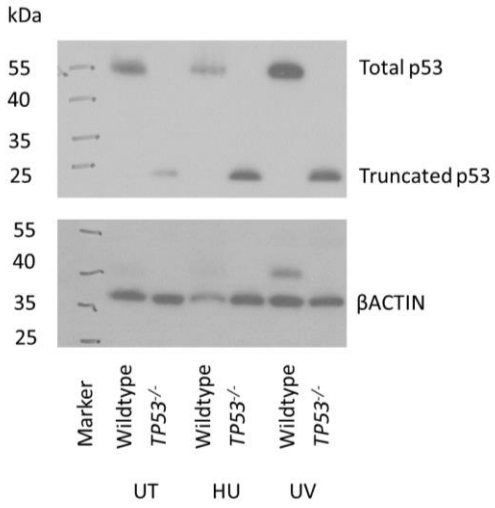
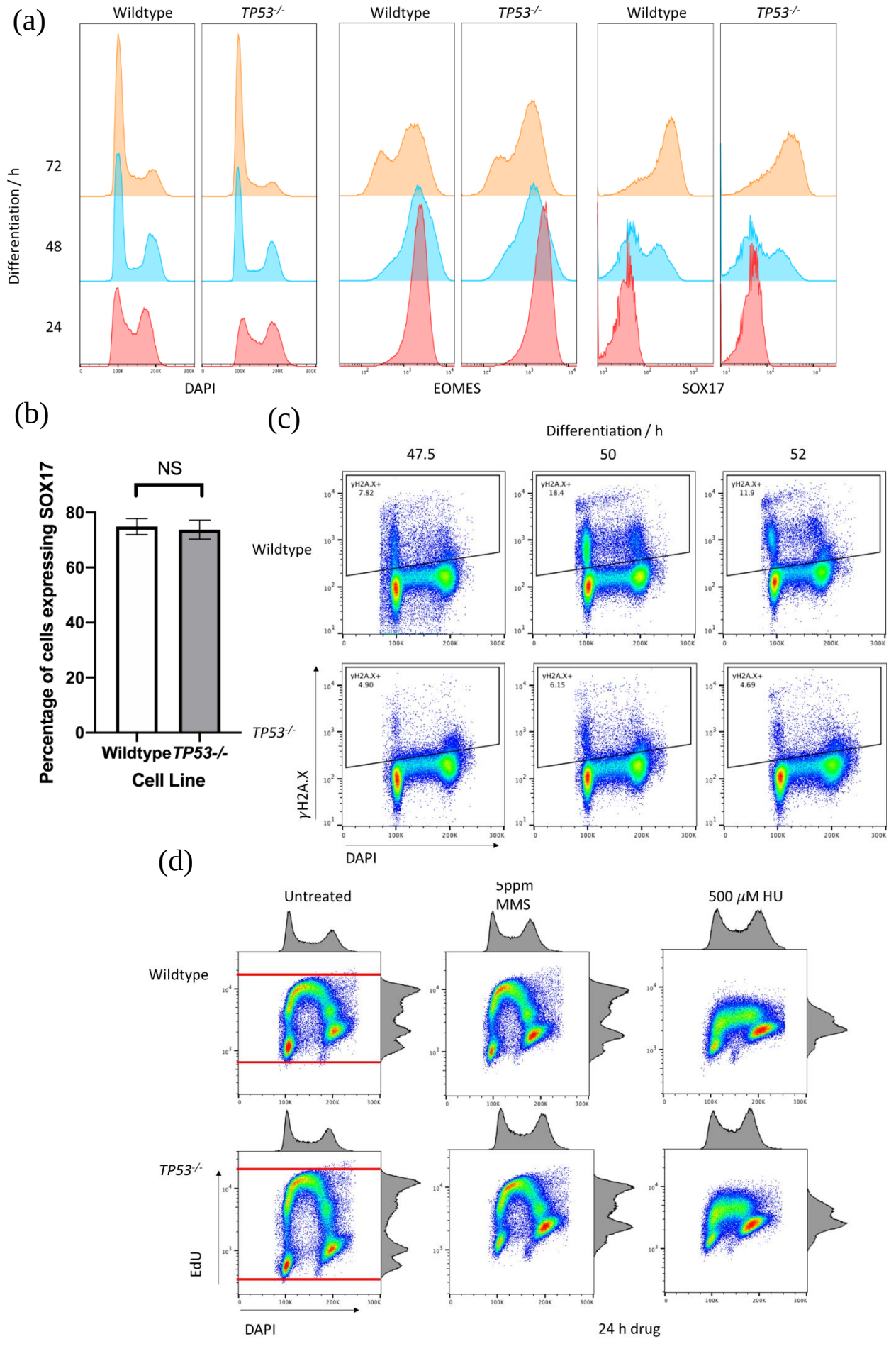


Figure 37. Analysis of the *TP53*$^{-/-}$ cell line using western blotting, βACTIN is shown as a loading control. The lack of the full length p53 is seen in the genetic knockout cell line.

TP53$^{-/-}$ definitive endoderm differentiation was monitored using permeabilised flow cytometry (Figure 38a), replicate experiments are shown in Figure 38b. This shows that knocking out *TP53* did not significantly affect the unperturbed differentiation efficiency. There was also a lower level of cell death during differentiation in these cells such that they were more confluent compared to the wildtype at 72 hours. This may suggest that they respond to external queues differently during differentiation, likely to be due to the role of p53 in apoptosis, which is discussed in Section 4.2.4.2.

TP53$^{-/-}$ cells were differentiated and collected for flow cytometry between 47 and 52 hours to analyse the γH2A.X spike (Figure 38c). Surprisingly, this data shows that the level of γH2A.X was lower at all timepoints compared to the wildtype cell line. Either the spike of upregulation associated with the EMT was missed and occurred at a different time in these cells, or the lack of p53 was causing a lower level of γH2A.X.

Figure 38. (See next page). Analysis of the *TP53*$^{-/-}$ cell line (a) SOX17, EOMES and DNA content at 24, 48 and 72 hours (b) replicate experiments showing the differentiation of the *TP53*$^{-/-}$ cell line compared to wildtype, n=6, p=0.8126 not significant, mean±SEM are plotted (c) the spike of H2A.X phosphorylation during differentiation (d) 2D-cell cycle plots without treatment or after treatment with MMS or HU in the undifferentiated state for 24 hours. The red lines indicate the increased EdU incorporation in the *TP53*$^{-/-}$ cell line.


(a)
Wildtype
TP53-/-
Wildtype
TP53-/-
Wildtype
TP53-/-
Differentiation / h
72
48
24
DAPI
EOMES
SOX17
(b)
NS
Percentage of cells expressing SOX17
80
60
40
20
0
Wildtype
TP53-/-
Cell Line
(c)
Differentiation / h
47.5
50
52
Wildtype
TP53-/-
γH2A.X+ 7.82
γH2A.X+ 18.4
γH2A.X+ 11.9
γH2A.X+ 4.90
γH2A.X+ 6.15
γH2A.X+ 4.69
γH2A.X
DAPI
(d)
Untreated
5ppm MMS
500 μM HU
Wildtype
TP53-/-
EdU
DAPI
24 h drug

Interestingly, the cell cycle profile of *TP53*$^{-/-}$ cells differed from wildtype cells in that they seemed to synthesise DNA faster in S phase, as shown from the increased EdU fluorescence (Figure 38d). However, both the wildtype and *TP53*$^{-/-}$ cells reacted in a similar manner to MMS and HU treatment. The *TP53*$^{-/-}$ cells also had the late replication bulge in the EdU cell cycle plot suggesting that this checkpoint was activated even without p53.

The level of MDM2 expression during differentiation in both wildtype and *TP53*$^{-/-}$ cells was examined using western blotting (Figure 39). This figure shows that the level of MDM2 was high in the undifferentiated cells and the total level decreased throughout differentiation: no difference between the wildtype and *TP53*$^{-/-}$ cells was seen. This suggested that p53 and MDM2 were regulated in a similar manner to one another, and the levels of both decrease during differentiation. Since this was only performed once in the *TP53*$^{-/-}$ cell line, it cannot be ascertained whether the difference in MDM2 is different in the two cell lines.

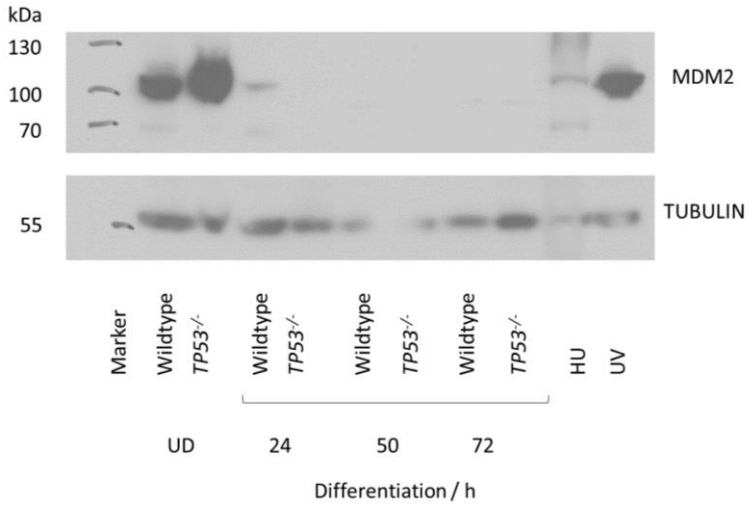


Figure 39. Analysis of MDM2 during differentiation using western blotting in the wildtype and *TP53*$^{-/-}$ cell lines. UD is the undifferentiated state. 2 mM HU and 20 J/m^2 UV were used as positive controls in the undifferentiated state, cells were treated six hours prior to collection.

4.2.2 Perturbing differentiation with low dose hydroxyurea

Cells were differentiated in the presence of HU and collected for permeabilised flow cytometry to assess whether this replication impediment could alter differentiation. This was first performed using 50 - 100 µM HU, which did not affect growth in the undifferentiated cells and this was then increased to 300 – 500 µM when no changes to differentiation were seen. Using 300 µM of HU or higher over the 72 hours of differentiation caused much greater cell death compared to the undifferentiated state, slowed the cell cycle, and increased the cell size which is indicative of a G2/M block. Cells were differentiated in the presence of 300 or 500 µM HU and collected every 24 hours (Figure 40a). This shows that the cells were not differentiating efficiently, and that the cell cycle was perturbed. EOMES was expressed in HU treated cells,

although possibly to a lower level than the untreated cells, and SOX17 was expressed in around 50% of cells. This suggested that HU was able to cause a change in the differentiation, however it is possible that the decrease in SOX17 positive cells was due to HU slowing the cell cycle. The correlation of EOMES and SOX17 expressing cells was monitored (Figure 40b). This suggested that most cells expressing SOX17 were also expressing EOMES, implying a dependency of EOMES on SOX17. This was more pronounced with HU treatment, perhaps because this was causing slowing of growth. However, since this was only performed once, the significance of this results cannot be checked.

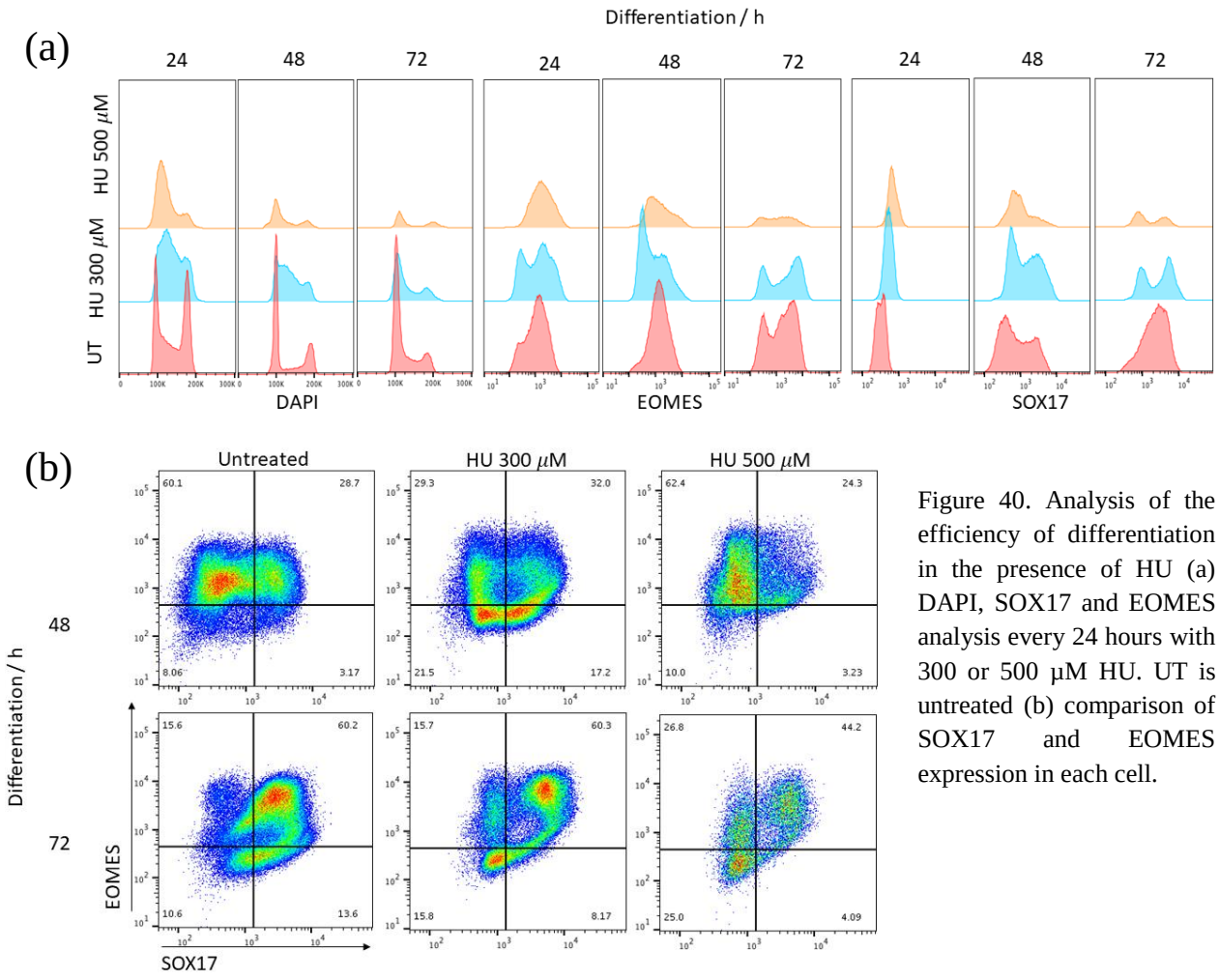


Figure 40. Analysis of the efficiency of differentiation in the presence of HU (a) DAPI, SOX17 and EOMES analysis every 24 hours with 300 or 500 μM HU. UT is untreated (b) comparison of SOX17 and EOMES expression in each cell.

4.2.3 Differentiation in the presence of DNA damaging agents

It is hypothesised that any replication fork blockage that causes the replicative DNA helicase to continue downstream of the obstruction to the replicative polymerase could cause epigenetic instability by the same means as the G4 at the BU-1 locus. In order to address this theory, while

also investigating how DNA damage affects the differentiation process, DNA damaging agents were added during differentiation.

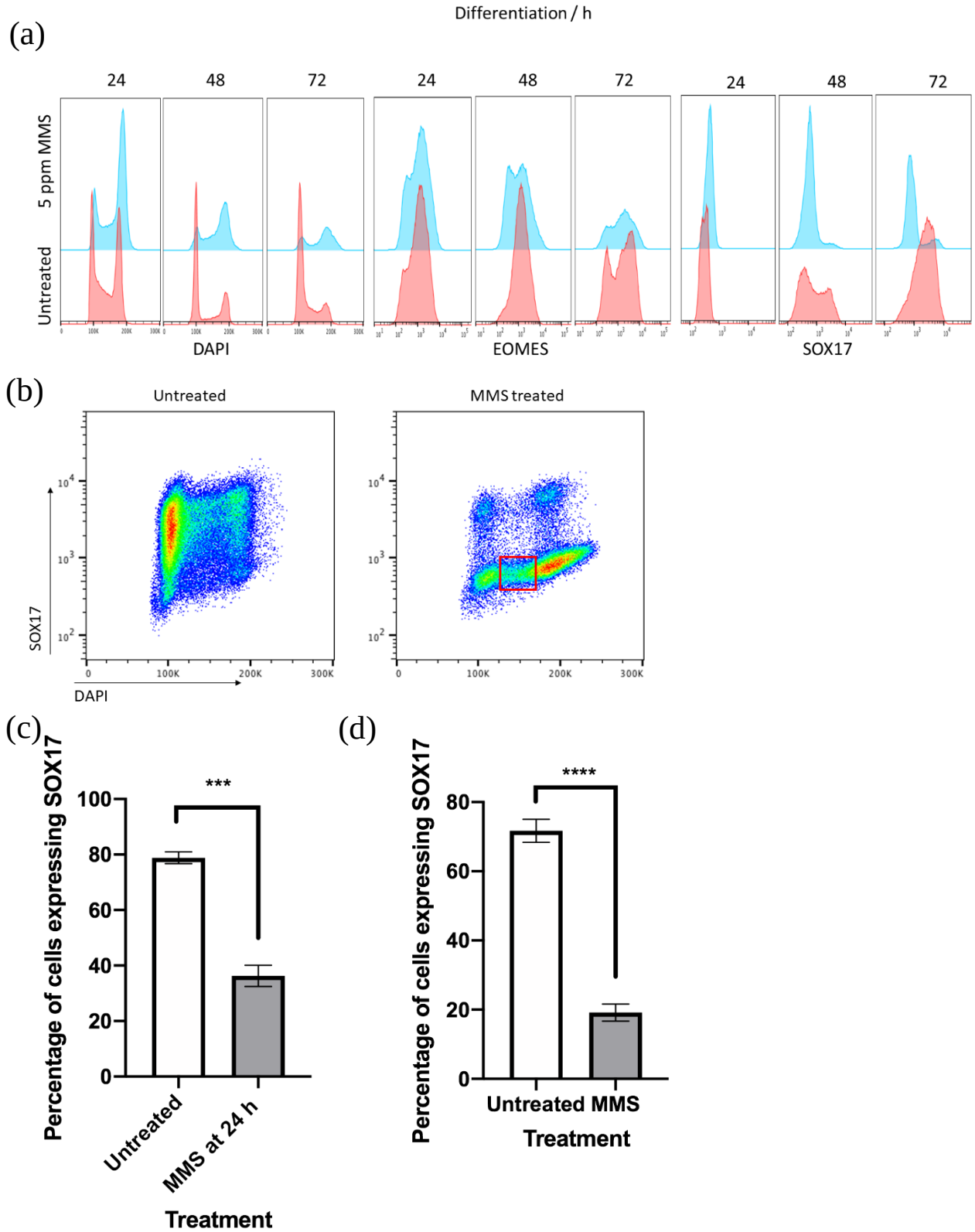
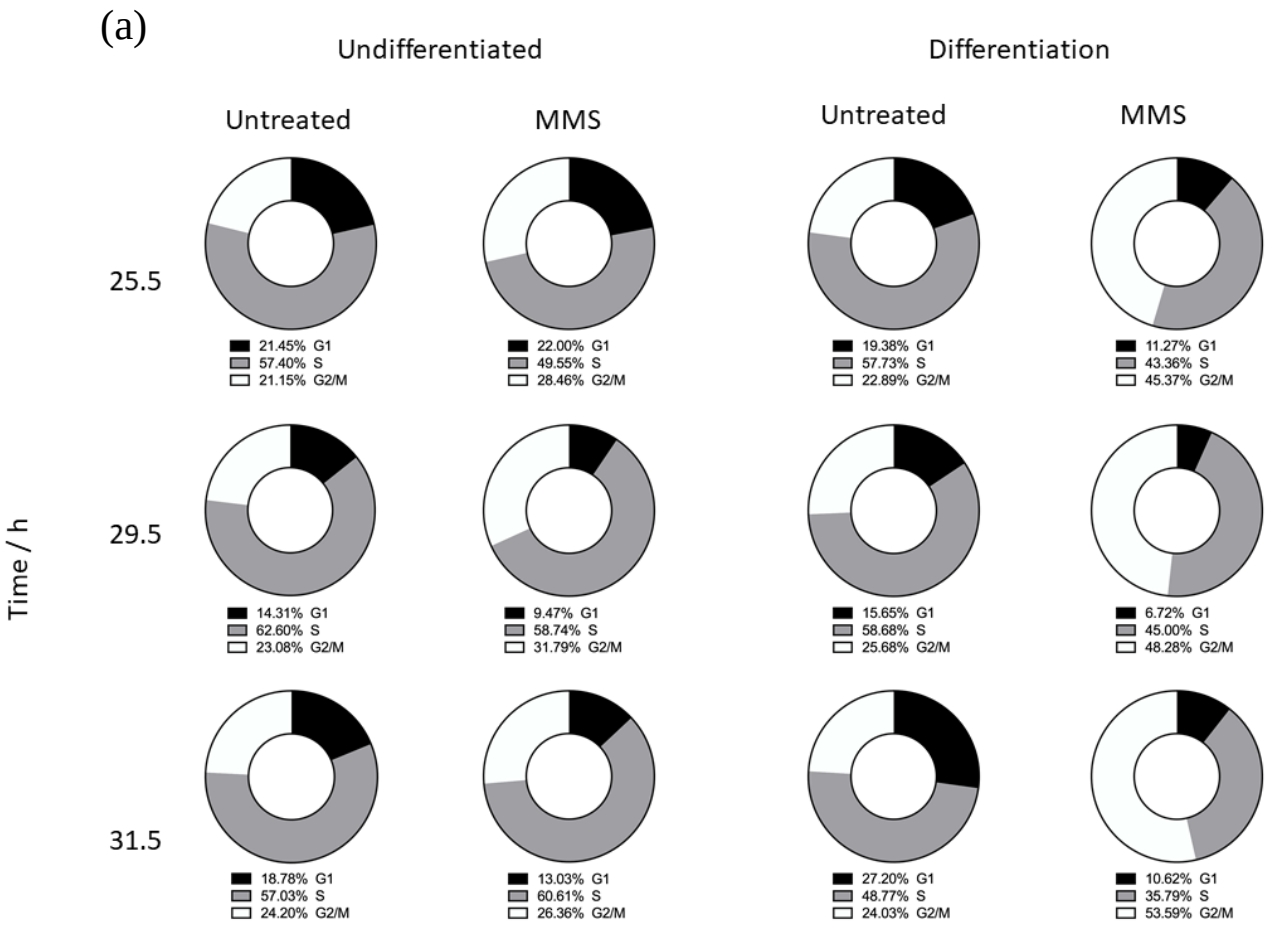


Figure 41. (See previous page) Flow cytometry analysis of wildtype cells treated with 5 ppm MMS at 0-, 24- and 48-hours (a) analysis of DNA content, EOMES and SOX17 (b) analysis of SOX17 at 72 hours and the proportion of the cells in S phase not expressing SOX17 (red box) (c) replicates of individual experiments n=8, where MMS was added at 0, 24 and 48 hours and the proportion of cells expressing SOX17 at 72 hours was compared to untreated samples, $p<0.0001$ using paired t test, mean±SEM are plotted (d) as for (c) but treating the cells only at 24 hours and cells were collected at 72 hours, n=7, $p=0.0001$.

4.2.3.1 Methyl methanesulfonate treatment during endoderm differentiation

MMS was well tolerated in the undifferentiated state; three doses of 5 ppm over 48 hours caused very little cell death or change in the cell cycle profile. Cells were differentiated with 5 ppm MMS added at 0-, 24- and 48-hours during differentiation; they were collected, and the differentiation efficiency was analysed using permeabilised flow cytometry (Figure 41a). It was evident that adding MMS during differentiation caused much more cell death compared to adding it in the undifferentiated state (Figure 24), suggesting that pluripotent cells were more able to deal with this treatment. The cell cycle profile suggested a similar pattern; more G2/M arrest occurred in the differentiated compared to the undifferentiated state. The addition of MMS in the early timepoints of differentiation caused a notable increase in the proportion of cells in G2/M, and there was no increase in the proportion of cells in G1 phase of the cell cycle at the endpoint of differentiation as there was in the untreated cells. The expression of EOMES was turned on as in the untreated, although may have been downregulated earlier. However, the most definite result was that SOX17 was not expressed at 48 hours and very few cells expressed this protein at 72 hours (Figure 41c and d). Of the cells in S phase of the cell cycle at 72 hours (Figure 41b), most of them were SOX17 negative, suggesting that this proportion of cells had not committed to differentiate and remaining in S phase was preventing the cells differentiating.

To look more closely at the changes in the cell cycle profile during the early stages of differentiation and to see how this compared to treatment in the undifferentiated state, cells were treated with 5 ppm MMS at 0- and 24-hours and EdU was added for an hour prior to collection. The proportion of cells in each phase of the cell cycle was quantified and the results are shown in Figure 42a. Flow cytometry plots at a range of time points show clearly that there was a high proportion of cells in the G2/M phase in the undifferentiated state in cells treated with MMS. However, this was dramatically increased during differentiation. This increase in G2/M at 24 hours of differentiation with MMS treatment is shown in Figure 42b. To understand whether the efficiency of differentiation was dependent upon the dose of MMS added, different concentrations were added 24 hours into the differentiation and cells were collected at 72 hours. The reason that 24 hours was chosen is explained in Section 4.2.5: 24 hours induced the greatest perturbation of differentiation. Figure 43a shows that the differentiation efficiency was related to the dose of MMS added between 2.5 and 10 ppm. However, there was an increase in γH2A.X in these cells as the dose of MMS increased (Figure 43b), suggesting that there was an increase in DNA damage and this may have prevented efficient differentiation by some means.



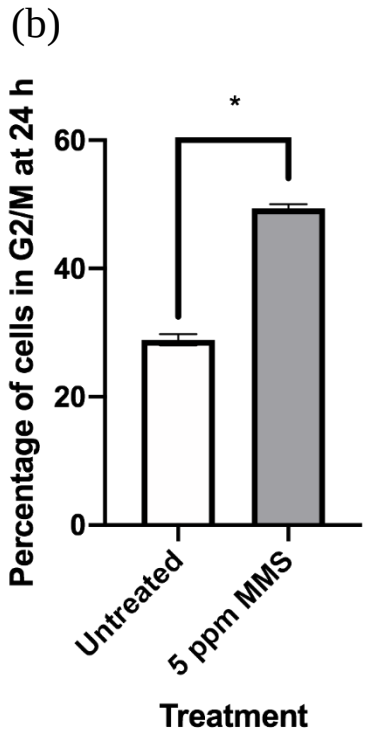


Figure 42. Cell cycle analysis after treatment of cells with MMS (a) EdU analysis of the proportion of cells in each phase of the cell cycle at 25.5, 29.5 and 31.5 hours in the undifferentiated state and during differentiation. G1 is shown in black, S in grey and G2/M in white. Cells were treated with 5 ppm MMS at 0- and 24-hours. (b) replicates during differentiation of cells treated with MMS at 0 hours and analysis at 24 hours, n=2 independent experiments, p=0.0482 using paired t test, mean±SEM plotted.

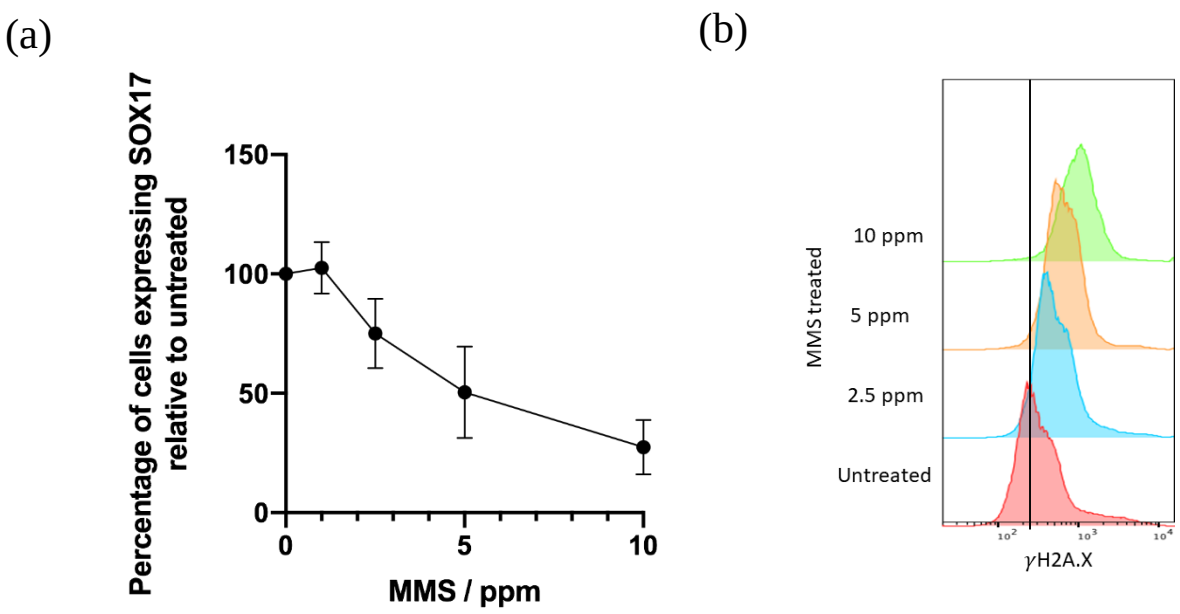


Figure 43. The dose response of MMS treatment at 24 hours to the outcome of differentiation (a) the percentage of SOX17 positive cells relative to untreated at 72 hours, n=2 independent experiments, mean±SEM is plotted (b) the phosphorylation of H2A.X in response to damage at 72 hours, the black line shows the untreated level in one experiment.

4.2.3.2 UV irradiation during endoderm differentiation

To address how UV damage impacts the differentiation process, a single dose of UV was added at 24 hours into the differentiation, the reason for which is discussed in Section 4.2.5. Cells were differentiated as usual but irradiated with 2 J/m^2 UV at 24 hours. While this had very little effect on the undifferentiated cells such that there was very little cell death (Figure 25), the differentiated cells showed an increased level of cell death. This cell death increased with increasing dose, as with MMS treatment during differentiation. Cells were collected for flow cytometry at 72 hours and the level of SOX17 was monitored (Figure 44a and b). Upon treatment with UV, a population of cells did not express SOX17 at 72 hours, suggesting that causing damage prevented the differentiation occurring as efficiently. As with HU treatment, there were also fewer cells expressing EOMES at 72 hours compared to untreated, also suggesting that differentiation is perturbed, possibly from an early stage. It is also interesting that there were few cells blocked in G2/M phase of the cycle, which was different to MMS treated cells.

In order to see if there was a differentiation efficiency dose response to UV treatment, different doses of UV were added at 24 hours and the cells were collected for flow cytometry at 72 hours (Figure 44b). At UV doses of 5 J/m^2 and above, there was a high level of cell death: the doses 0.5, 1 and 2 J/m^2 were analysed. The results show that there was a clear UV dose response in the proportion of cells expressing SOX17 at 72 hours, in a similar manner to MMS treatment.

However, since the dose response was only performed once due to the high level of error in using the UV box, this response is unlikely to be as significant as with the MMS treatment.

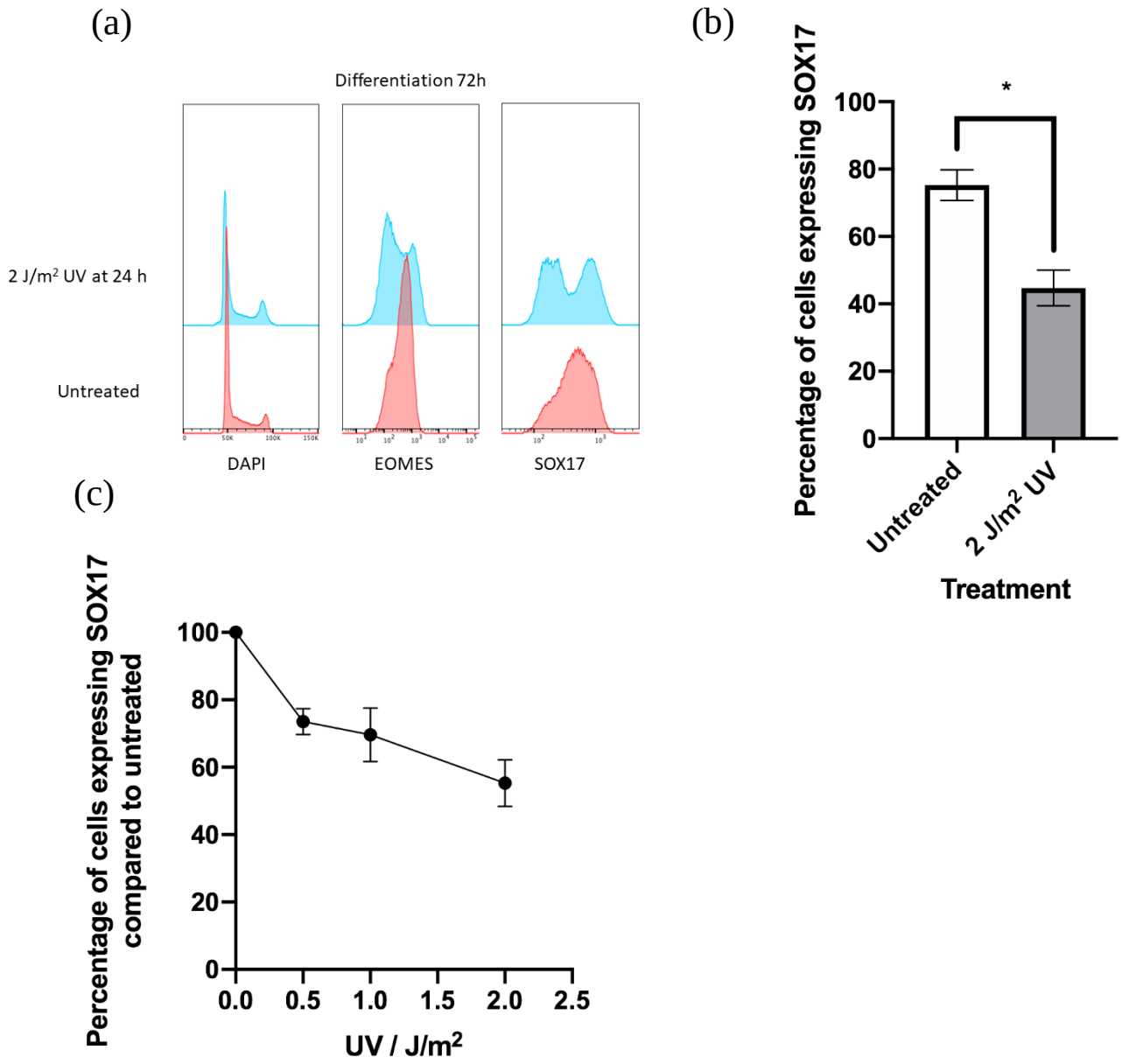


Figure 44. Analysis of the efficiency of endoderm differentiation with 2 J/m^2 UV irradiation at 24 hours (a) DAPI, EOMES and SOX17 at 72 hours (b) the percentage of cells expressing SOX17 at 72 hours after treatment with 2 J/m^2 UV at 24 hours, n=3, p=0.0252, paired t test, mean±SEM are plotted (c) the dose response of UV treatment added at 24 hours on the percentage of SOX17 positive cells at 72 hours relative to untreated, n=2, mean±SEM are plotted.

4.2.3.3 Kinetics of the DDR after MMS treatment

At this stage it was clear that undifferentiated and differentiating cells responded differently to damage and were also able to tolerate it differently to somatic cells. In order to analyse the DNA damage response following MMS treatment, wildtype cells were treated with 10 ppm MMS in both the undifferentiated and differentiated state at 24 hours, and cells were analysed

for γH2A.X signal (Figure 45). The major signal was in S phase and did not appear until four hours in both the undifferentiated and differentiating state, much later than seen in somatic cells, and remained upregulated until at least eight hours after the induction of damage (Rogakou et al., 1998). The MMS was not washed out and the length of time it is active in aqueous solution is debated, so this could be a cumulative effect but is likely to represent the higher tolerance of hiPSCs to damage and shows their difference compared to somatic cells. There was also a similar change in the level of γH2A.X in both the undifferentiated and differentiating samples.

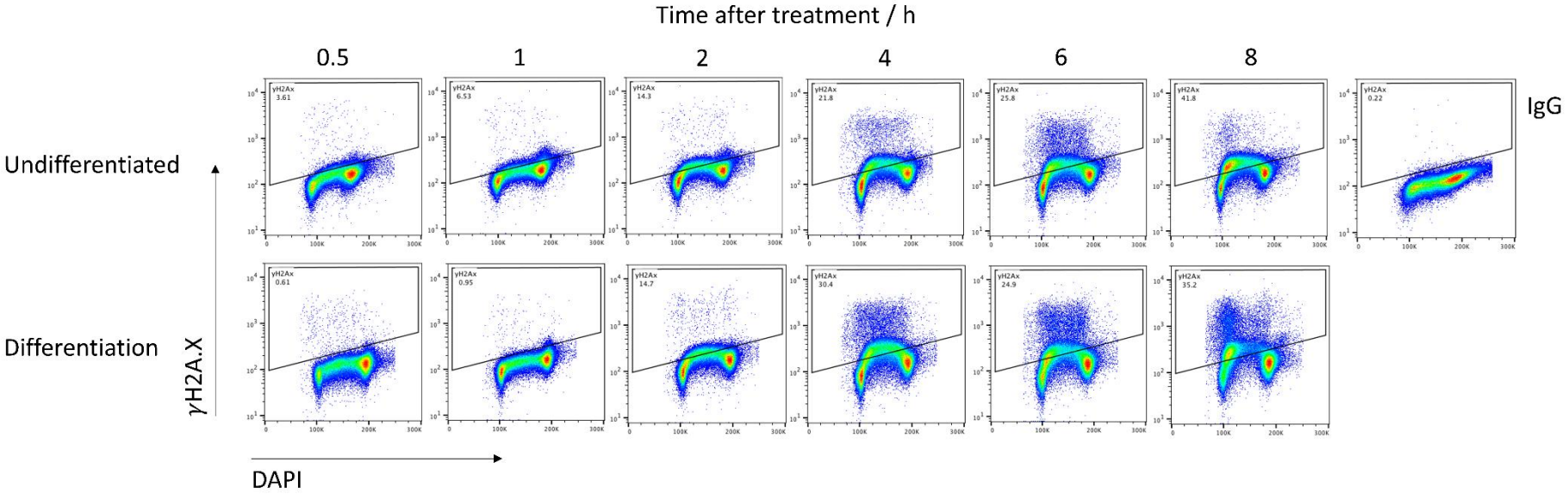


Figure 45. The phosphorylation of H2A.X after treatment with 10 ppm MMS monitored after set times in undifferentiated and differentiating cells. Cells were treated at 24 hours.

4.2.4 The DDR during perturbed endoderm differentiation

4.2.4.1 The γH2A.X spike after treatment with DNA damaging agents

To understand whether perturbing differentiation in the face of DNA damage could alter the γH2A.X that had been seen at around 50 hours during the normal differentiation process, cells were differentiated with 300 µM HU or were treated with 5 ppm MMS and the cells were collected for flow cytometry analysis (Figure 46). Both HU and MMS treatment were seen to cause an upregulation of γH2A.X in the undifferentiated state. During differentiation at the point where γH2A.X spike was seen, the signal was also upregulated in the HU and MMS treated cells but not to the same extent as in the untreated. The base level was also increased during differentiation, to a slightly higher extent than the undifferentiated treated cells showing the general DDR. The differentiated cells showed a higher γH2A.X signal in response to damage and therefore may have been tolerating the damage in this differentiated state less well than in the undifferentiated state. However, the high signal seen during this point in differentiation was clearly separate from the DDR.

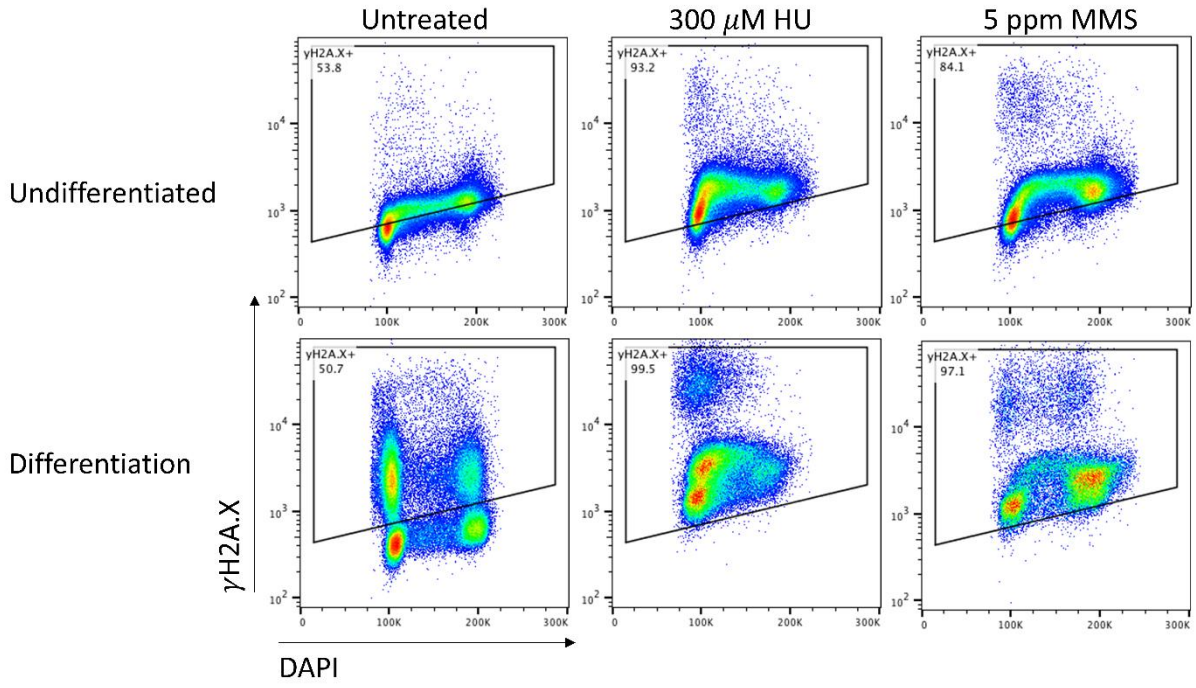


Figure 46. Phosphorylation of H2A.X at the spike that occurs in unperturbed endoderm differentiation. Cells were collected at 51.5 hours and monitored for phosphorylation. MMS treated cells were treated at 0-, 24- and 48-hours with 5 ppm, HU treated cells were cultured with 300 µM HU throughout differentiation. The baseline level of H2A.X phosphorylation increases in damaged cells, but the gating is in relation to the untreated differentiating cells to compare the level of phosphorylation in the undifferentiated state and during the spike.

4.2.4.2 The role of p53 during differentiation in the presence of DNA damage

In Section 4.2.1.1 it was shown that MMS and HU treatment during differentiation upregulated the level of p53 protein in the cells. This upregulation of p53 could be preventing the differentiation in these treated cells and if this were the case then a genetic knockout of *TP53* should prevent this. Previous data has shown that total levels of p53 decrease during the differentiation (Chang et al., 2011). The *TP53*$^{-/-}$ cell line, introduced in Section 4.2.1.1, was used in parallel to the wildtype hiPS cells, and differentiated with and without 5 ppm MMS added at 0-, 24- and 48-hours (Figure 47a). The *TP53*$^{-/-}$ cells exhibited a much lower level of cell death when treated with MMS compared to wildtype cells, suggesting they were not as sensitive to DNA damage or less able to undergo apoptosis due to the lack of p53. The treatment with MMS during differentiation in both cell lines caused an increase in the G2/M population of cells, suggesting that MMS caused a replication impediment in the *TP53*$^{-/-}$ cells as well. Interestingly however, the *TP53*$^{-/-}$ cells behaved in a very different manner to the wildtype cells: they differentiated to the same extent as the wildtype untreated cells (Figure 47b and c). This result clearly showed that the G2/M block was not preventing differentiation, but the presence of p53 was. Thus the cells did not have to be in G1 phase for efficient differentiation to occur.

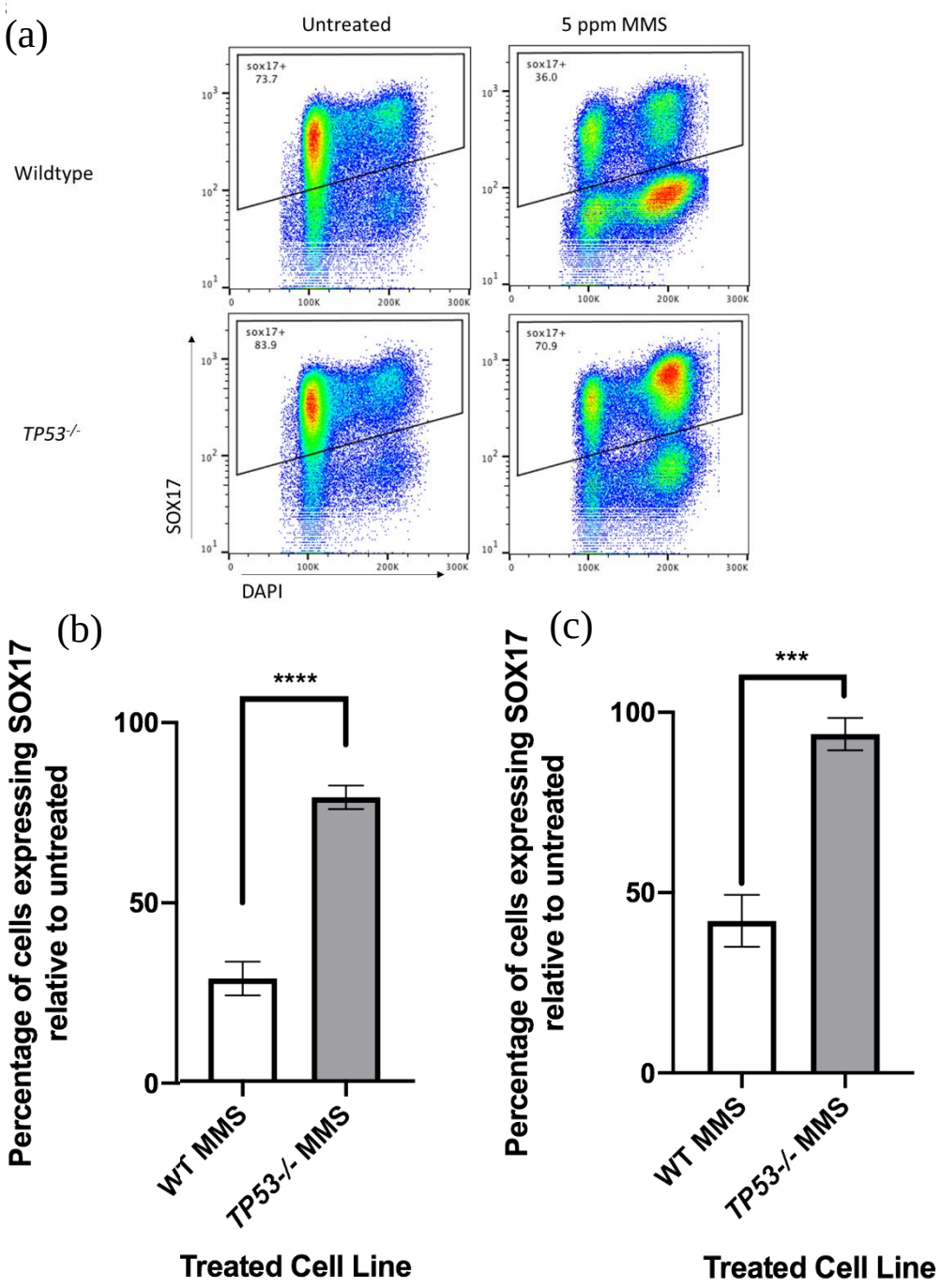


Figure 47. Wildtype and *TP53*$^{-/-}$ cells were treated with 5 ppm MMS every 24 hours of differentiation and collected at 72 hours to analyse the proportion of SOX17 positive cells and the phase in the cell cycle compared to untreated cells. (a) flow cytometry analysis of one experiment (b) replicate experiments on treatment with 5 ppm MMS at 0, 24 and 48 hours n=6, $p<0.0001$, unpaired t test, mean±SEM is plotted (c) as (b) but with treatment just at 24 hours n=4, $p=0.0009$.

UV irradiation was also used to induce damage in *TP53*$^{-/-}$ cells and compared to the wildtype cells as above with MMS treatment (Figure 48). The results were very similar: UV treatment caused a large G2/M blockage with 5 J/m^2 UV, specifically in the *TP53*$^{-/-}$ cell line, but the cells differentiated with a high efficiency in the *TP53*$^{-/-}$ cell line. This result was in line with the MMS treatment during differentiation.

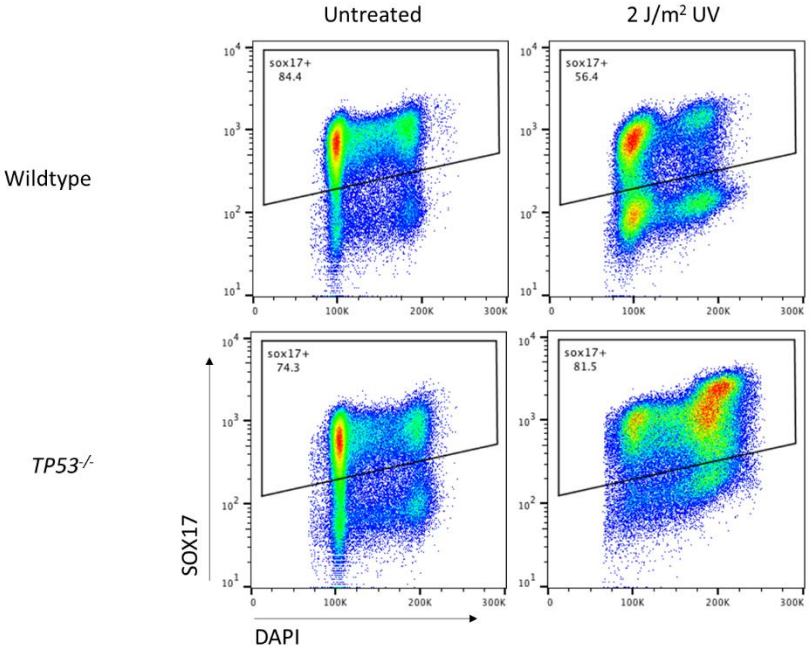


Figure 48. As in Figure 47 but with 2 J/m^2 UV irradiation at 24 hours.

It was likely that ATM would be phosphorylating both H2A.X and CHK2 in this system, and that having high levels of DNA damage would upregulate total CHK2 protein, allowing it to be phosphorylated around the time of the H2A.X phosphorylation. To attempt to prevent the large proportion of cells in the G2/M phase of the cell cycle when differentiating in the presence of MMS or UV, a CHK2 inhibitor was used (NSC109555 Tocris). However, the CHK2 inhibitor did not stop the G2/M blockage in MMS treated cells at 0.5 µM: the highest dose that did not cause toxicity to the cells (Figure 49). Higher doses prevented efficient differentiation and this strategy was not continued further.

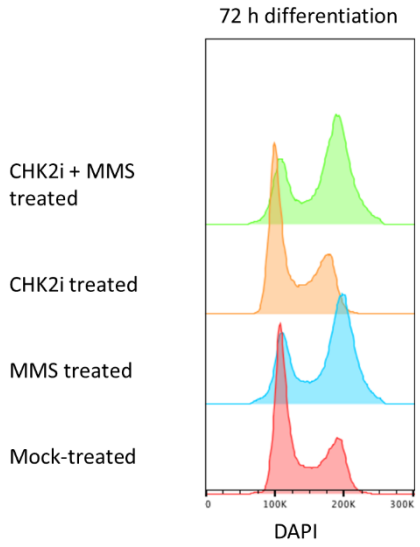


Figure 49. Flow cytometry analysis of the G2/M blockage induced in differentiation with cells treated with MMS every 24 hours. The CHK2 inhibitor was added at 0.5 µM in an attempt to prevent this blockage.

On the other hand, *ATM*$^{-/-}$ cells responded in an opposite manner compared to *TP53*$^{-/-}$ cells when DNA damaging agents were added during the differentiation process: the addition of MMS

caused the cells to become very sick and there was increased cell death (Figure 50). The differentiation efficiency was decreased dramatically compared to the untreated sample: this suggested that inflicting damage without ATM prevented cells differentiating efficiently, and that this protein may be required to deal with damage in differentiation. This experiment was only performed once due to the high level of cell death.

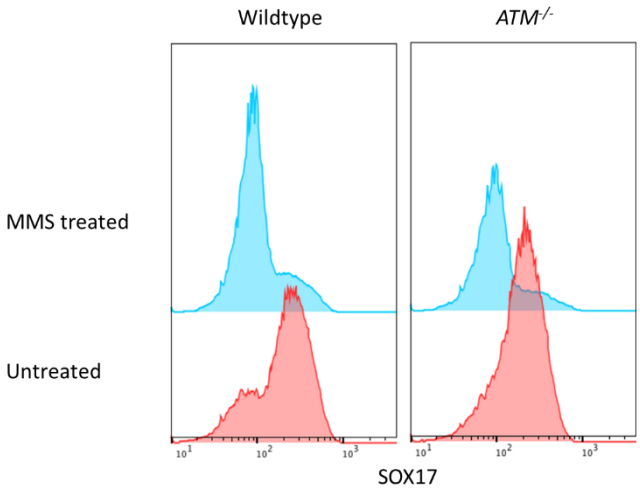


Figure 50. Wildtype and *ATM*$^{-/-}$ cells were treated with 5 ppm every 24 hours during differentiation and flow cytometry analysis was performed to at 72 hours to monitor the SOX17 positive cells.

4.2.5 Altering the time of DNA damage induction during differentiation

It was of interest to work out whether the time at which DNA damage was introduced could affect the outcome of the differentiation. To get an insight into whether there was a crucial window of time in which adding damage could affect the differentiation, single doses of 5 ppm MMS were added at different points (Figure 51a and b). This shows that differentiation was most affected when MMS was added at around 24 hours into the process. This time broadly correlates with the point in which the levels of DDR proteins decrease during differentiation. The resulting upregulation of DDR proteins after treatment with DNA damaging agents may prevent further differentiation in these cells, in effect a differentiation checkpoint. This window of time may also be crucial for cell fate choice and lineage commitment pathway decisions. The outcome of treatment with damaging agents may be affected by the position of the cells in the cell cycle when damage was induced, as the cell cycle distribution was altered throughout differentiation.

Cells were also treated with MMS after 72 hours to see whether this decreased the percentage of cells expressing SOX17. Cells were treated at different times after 72 hours and collected at 79 hours (Figure 51c). There was no decrease in SOX17 expression after treatment with MMS, suggesting that MMS had to be added during the process of differentiation to affect the outcome. However, collecting the cells at a later point after MMS addition could have increased the effect.

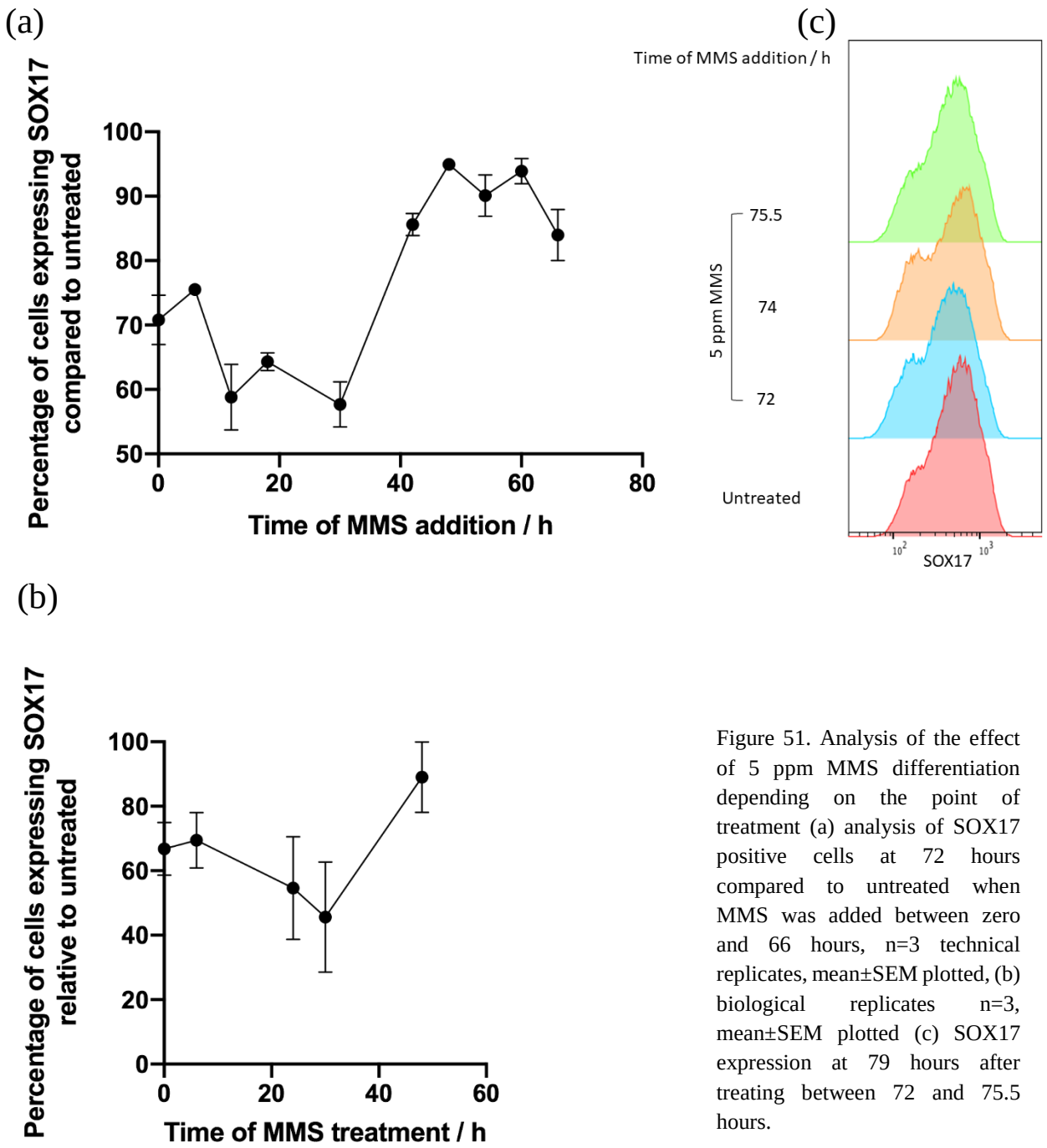


Figure 51. Analysis of the effect of 5 ppm MMS differentiation depending on the point of treatment (a) analysis of SOX17 positive cells at 72 hours compared to untreated when MMS was added between zero and 66 hours, n=3 technical replicates, mean±SEM plotted, (b) biological replicates n=3, mean±SEM plotted (c) SOX17 expression at 79 hours after treating between 72 and 75.5 hours.

4.2.6 Does inflicting DNA damage prevent cell cycle progression?

A simple answer to why inducing DNA damage was preventing differentiation was that it was either slowing or completely stopping the cell cycle and therefore the cells were not becoming specified to definitive endoderm.

4.2.6.1 Does the length of the cell cycle increase in cells treated with DNA damage inducing agents?

If the damage exposed cells were not differentiating because they were lagging in the number of cell divisions compared to untreated cells, then letting these cells differentiate for longer should increase the efficiency of differentiation to untreated levels. Cells were treated with 5 ppm MMS at 0-, 24- or 48-hours and the cells were collected for flow cytometry at 96 hours to allow an extra 24 hours of differentiation after treatment (Figure 52). This result shows that the extra 24 hours did not allow the treated cells to catch up with the untreated differentiation and suggests that the undifferentiated cells at 72 hours do not go on to differentiate given an extra 24 hours. Maintaining the untreated cells in this state for over 118 hours caused the cells to become too confluent and lose adherence so timepoints later than 118 hours were not used. The percentage of SOX17 in untreated samples at later timepoints also decreased.

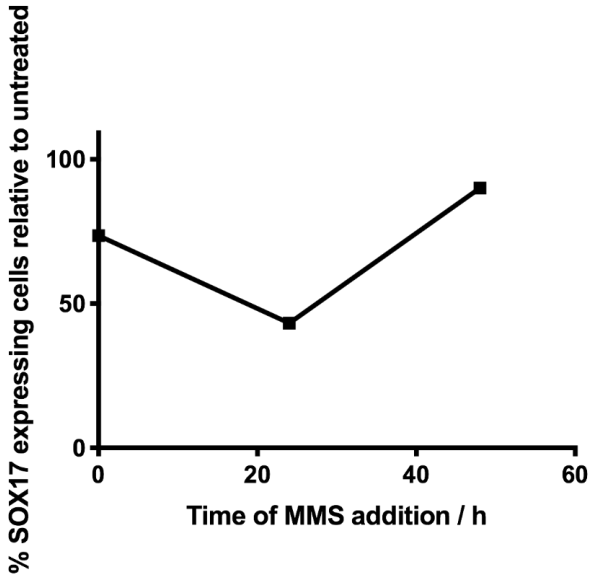


Figure 52. The percentage of SOX17 positive cells at 96 hours (after an extra 24 hours of differentiation) after treatment with 5 ppm MMS at 0-, 24- or 48-hours. There was a decrease in expression regardless of the time of treatment, and treatment at 24 hours still caused the biggest decrease. One experiment shown.

4.2.6.2 Does the cell cycle stop in damaged cells?

In order to understand whether cells treated with DNA damaging agents were not differentiating efficiently simply because they had stopped dividing, the number of cell divisions was monitored. This was done using the CellTrace™ Cell Proliferation Kit (ThermoFisher) which

stains the cells with a fluorescent dye. The fluorescence of a cell decreases by half at every cell division, allowing the number of cell divisions that a given cell has gone through to be counted. This method was performed in the undifferentiated and differentiated state, with and without the addition of 5 ppm MMS at 24 hours and the cells were collected at different timepoints and analysed using flow cytometry (Figure 53a). The cell division staining showed that cells treated with MMS during differentiation did have a small but significantly different decrease in the number of cell divisions (Figure 53b). However, SOX17 was able to be expressed from cell cycle number three in the untreated cells. The slight cell cycle slowing of MMS treated cells was also the case in *TP53*$^{-/-}$ cells (Figure 53c and d), showing that the lack of p53 did not increase the percentage of SOX17 positive cells by simply increasing the cell cycle speed. However, the experiment in the *TP53*$^{-/-}$ cell line was only performed once and would need to be verified. This data suggested that it was not a decrease in the number of cell cycles that was inhibiting efficient differentiation.

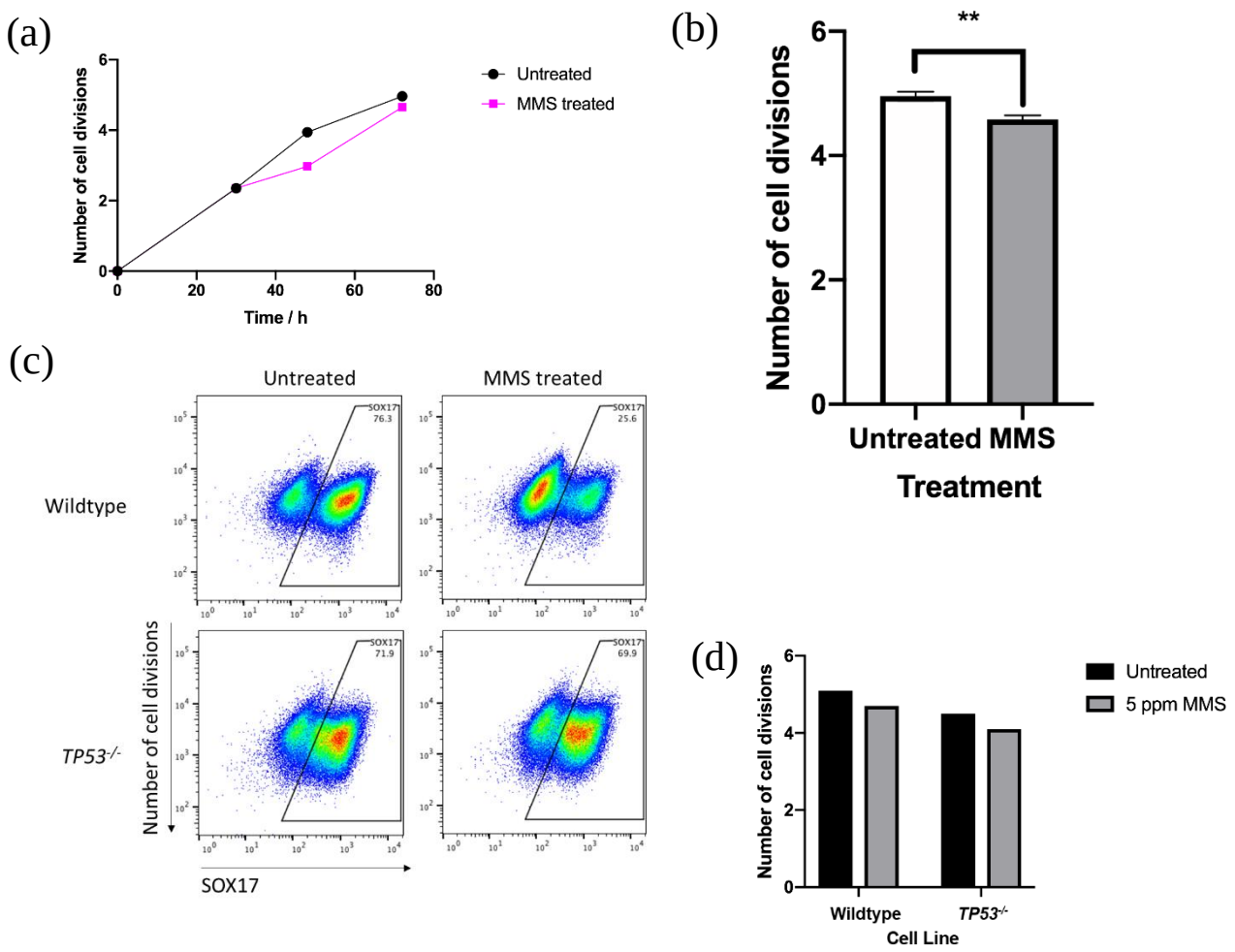


Figure 53. CellTrace™ analysis of the number of cell divisions occurring during differentiation (a) 5 ppm MMS was added at 24 hours and cells were collected for analysis at 24-, 48- and 72-hours. Untreated cells had gone through a median of 4.96 cell cycles compared to the MMS treated cells which had done 4.65. (b) replicate experiments of the CellTrace™ experiment showing the number of cell cycles at 72 hours will MMS treatment at 24 hours compared to untreated cells. Mean±SEM is shown, n=3, p=0.0097 use paired t test (c) flow cytometry to show the number of cell divisions with *TP53*$^{-/-}$ fluorescence decreases by half on each cell division (d) a graph showing the number of cell divisions from (d).

4.3 Cell death during differentiation

During the unperturbed differentiation *in vitro* there was consistently large amounts of cell death compared to the undifferentiated state, an increased number of cells lost adherence throughout the process and floated to the surface compared to the undifferentiated state. To understand why this might be happening, Annexin V-7AAD cell death staining was performed (Figure 54a) and analysed using flow cytometry (Figure 54b).

4.3.1 Analysis of the type of cell death during endoderm differentiation

Surprisingly this showed that at 24 and 48 hours into the differentiation process, the cells that had lost adherence and were in the supernatant were mostly alive, with a proportion of late

apoptotic cells. This suggests that these cells were not dying and therefore may have been losing adherence for another reason, such as not differentiating correctly. At 72 hours most of the supernatant cells were late apoptotic suggesting that programmed cell death was occurring again at the end of the differentiation process. However, during the differentiation process, the adherent cells seem to be showing increased apoptosis at the later time points. This shows that there was a large amount of cell death *in vitro* in the differentiation process although the reasons why remain unknown. In the undifferentiated state, the cells that had lost adherence were mostly apoptotic cells, and therefore not alive. There were also much fewer 'floating cells' in the undifferentiated cells compared to during differentiation. This experiment shows comparable data to the mES 'floating cell' experiments and shows that a large amount of data can be obtained from the cells that lose adherence.

(a)

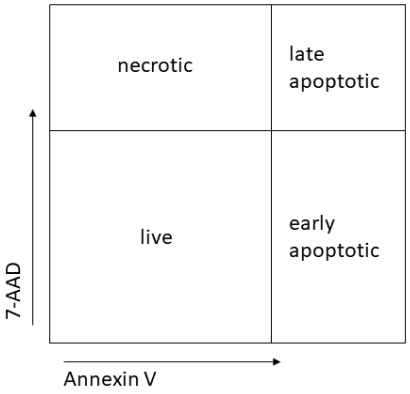
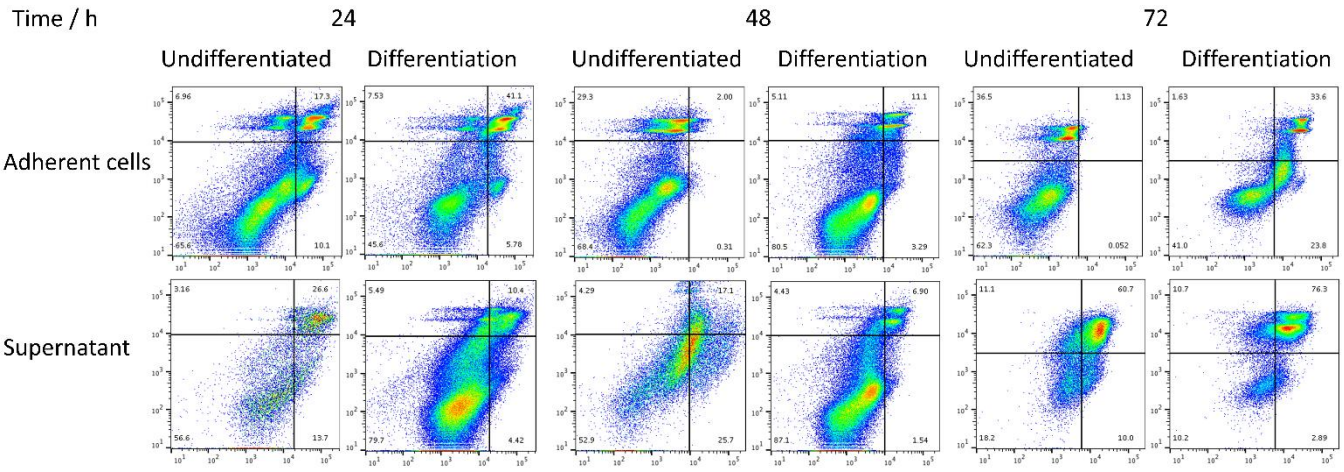


Figure 54. Annexin V / 7-AAD staining to analyse necrotic and apoptotic cells (a) schematic to explain the gating (b) untreated wildtype undifferentiated and differentiating cells were collected every 24 hours. The adherent and non-adherent population were separated prior to analysis.

(b)



4.3.2 Analysis of the reason behind losing adherence during differentiation

To see whether the cells that had lost adherence during the differentiation process were expressing either EOMES or SOX17, adherent and non-adherent cells were collected at 72 hours and permeabilised flow cytometry was performed to check the protein levels (Figure

55a). The cells that had lost adherence displayed very low SOX17 or EOMES, suggesting that not expressing the markers could prevent them from remaining adherent in the culture conditions. However, it could be that the cells were initially losing adherence and this loss of adherence was preventing the expression of the proteins EOMES and SOX17. It was also clear that the non-adherent cells were not cycling and did not show a cell cycle profile typical of differentiated cells.

To check which proteins were expressed only in the alive cells in the supernatant, non-adherent cells were collected and the live cells were separated from dead cells using the MACS® dead cell removal kit. These live cells were fixed and analysed for the expression of EOMES and SOX17 (Figure 55b). The non-adherent live cells at 24 hours were not expressing EOMES. This suggested that the reason that they were no longer attached to the plate was because they were not differentiating towards the endoderm lineage. At 48 hours into the differentiation none of the alive cells were expressing SOX17 when it would be expected that around 40-60% of differentiating cells would be expressing it at this time. However, they were expressing EOMES at this time and therefore had partially differentiated. This implies that at least some of the cells that did not remain adherent during the differentiation were not dying and were also not differentiating. This could be caused by a problem occurring during differentiation, after EOMES was switched on but before definitive endoderm was specified.

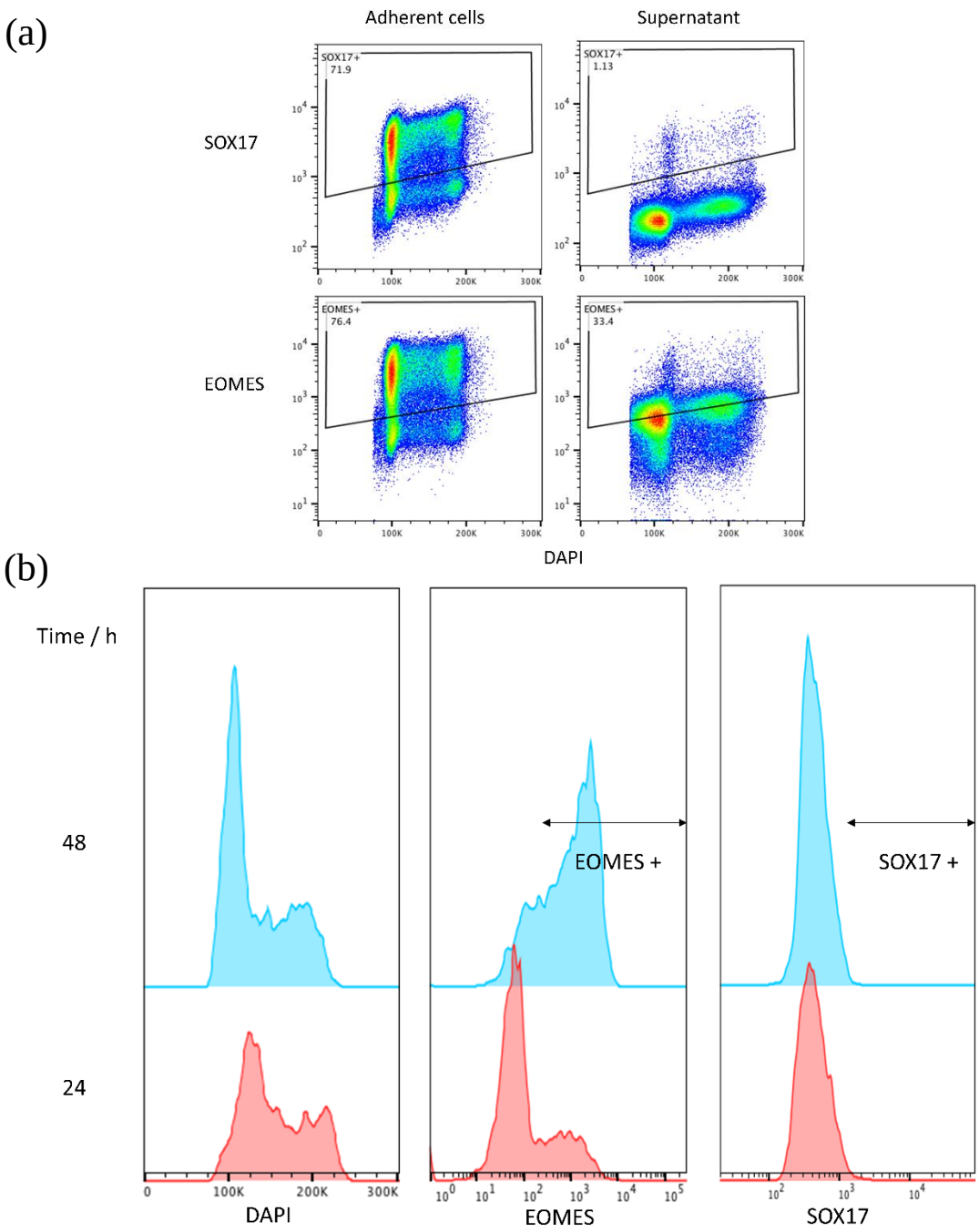


Figure 55. Analysis of the expression of non-adherent cells (a) the expression of SOX17 and EOMES at 72 hours in the 'floating cells' (b) the gene expression of live cells in the supernatant at 24 and 48 hours, the EOMES and SOX17 positive population in adherent cells is marked with an arrow.

4.4 Using RNA sequencing to understand the consequence of inflicting DNA damage on differentiation

As showed throughout this chapter, treatment of cells with DNA damaging agents caused a very different effect during differentiation compared to in the undifferentiated state. The levels of DDR proteins, the quantity of cell death and the DNA damage response upon treatment with MMS, UV and HU were very different between these states. The *TP53*$^{-/-}$ cell line was also able to differentiate to a greater extent when treated with DNA damaging agents compared to wildtype cells. In order to understand the differences seen throughout the differentiation, RNA sequencing was performed in wildtype and *TP53*$^{-/-}$ cells, in both the undifferentiated state and during differentiation, with and without MMS treatment. MMS was added to cells at 5 ppm at 24 hours and the cells were collected for RNA every 24 hours in triplicate.

The principle component analysis (PCA) plot for the total experiment is shown in Figure 56a; this gave an overview of the data and showed the 'vector of differentiation'. The basis of a PCA is to simplify the data to, in this instance, two dimensions to display the greatest variance between the individual datasets. This allows a pattern in the data to be seen before analysing individual genes. This showed that the undifferentiated samples clustered well together, as did the 24-, 48- and 72-hour samples. During differentiation, before MMS treatment, at 24 hours the wildtype and *TP53*$^{-/-}$ cells were defined in slightly different locations but close to one another. At 48 hours the wildtype and *TP53*$^{-/-}$ triplicates remained very close but the wildtype MMS treated did separate out slightly from the wildtype untreated: this is seen in a similar manner in the PDS treated cells in Results III. This was not the case for the *TP53*$^{-/-}$ cells. At 72 hours there was a definite clustering of wildtype and *TP53*$^{-/-}$separately with the wildtype MMS treated and untreated forming separate clusters. The *TP53*$^{-/-}$ treated and untreated did not separate as well, possibly due to the MMS having less effect in this cell line.

Although the undifferentiated cells did cluster together there were clear gene expression changes between these cells and the PCA was rerun with just the undifferentiated cells and is shown on a different scale due to the number of deregulated genes being much smaller (Figure 56b). This showed that the wildtype cells all clustered to the left of the plot and the *TP53*$^{-/-}$ to the right of the plot. The 24-, 48- and 72-hour RNA each clustered together, showing that there was a 'day effect' discussed further in Results III and Results IV. The untreated and wildtype MMS treated clustered separately, which was less pronounced with *TP53*$^{-/-}$, again highlighting

the fact that the wildtype cells react transcriptionally in a different manner to *TP53*$^{-/-}$ cells when treated with damage.

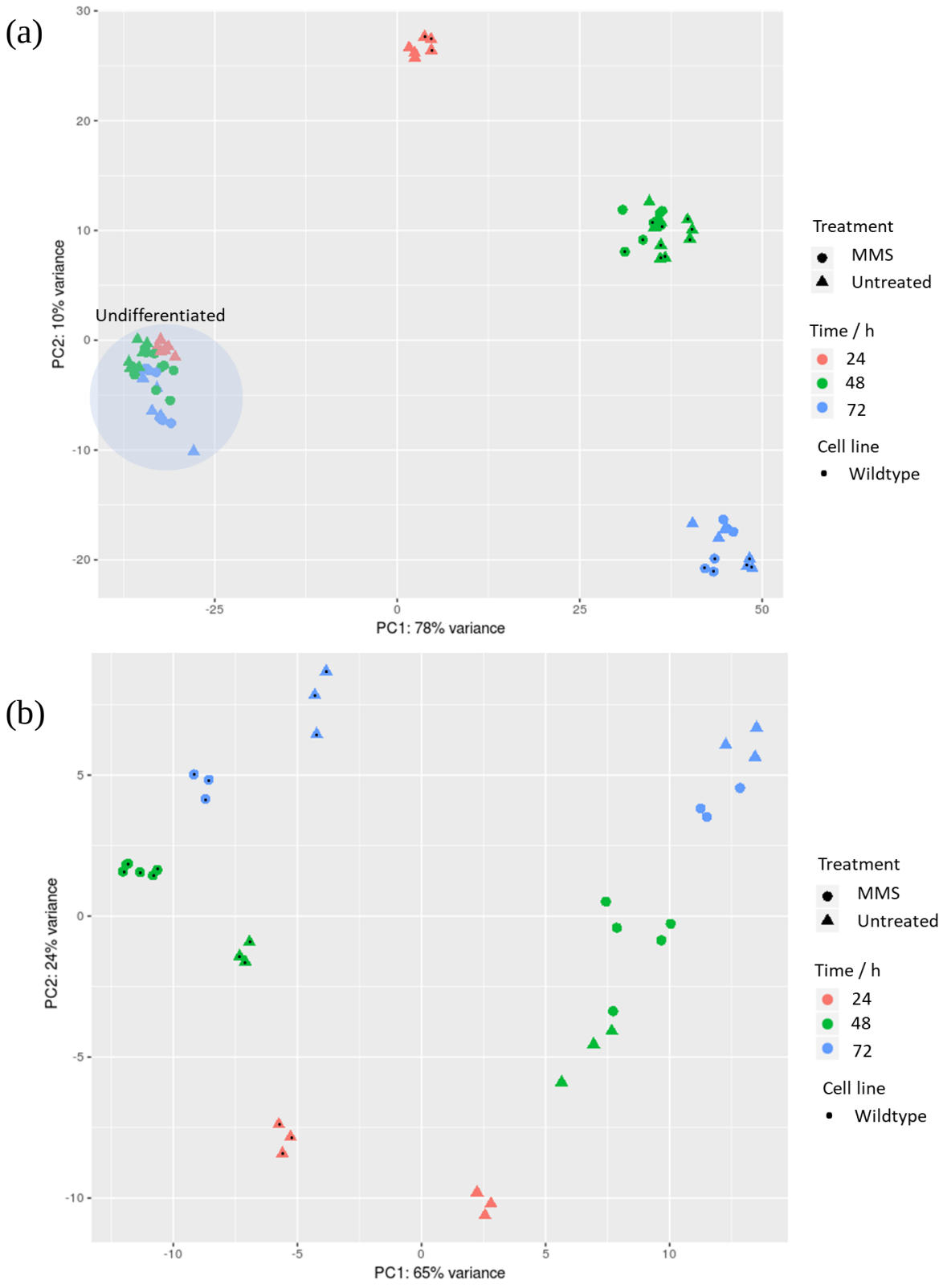


Figure 56. PCA analysis of this dataset to show the variance of the datapoints (a) the undifferentiated points are circled and the *TP53*$^{-/-}$ and wildtype are not marked due to the similarity on this scale. Cells were treated with 5 ppm MMS at 24 hours (circles) or untreated (triangles), 24 hours (red), 48 hours (green) and 72 hours (blue), wildtype cells have a central black dot whereas *TP53*$^{-/-}$ is left as symbols. Data was performed in triplicate and libraries at 48 hours of differentiation were run twice for sufficient reads (hence double the points) (b) as above but just with undifferentiated samples.

4.4.1 Changes to the expression of DDR markers during unperturbed differentiation

As the western blots showed changes in DDR markers during differentiation, the RNA of these transcripts was checked and is shown in Figure 57. The RNA used for these graphs were taken from the RNAseq experiment in Results III as this was performed first. Interestingly there were a number of genes with altered expression throughout differentiation, but there were also a number where there were not significant differences despite knowledge that there is a difference in the protein level. For example, *TP53* mRNA did decrease slightly during differentiation but not significantly and it is known that MDM2 controls the level of p53 protein during differentiation and is expressed to a similar level. *CHEK2* mRNA decreased significantly during differentiation, in line with the idea that it may decrease so that it cannot be phosphorylated by ATM at the EMT. *RPA2* decreased significantly at 48 hours, also at the time of the EMT, and this RPA subunit is able to be phosphorylated by ATM. *PARP1* and *TERT*, a subunit of telomerase, decreased throughout differentiation in line with these being expressed at higher levels in the undifferentiated cells. Interestingly the level of *TP53BP1* increased throughout differentiation, potentially suggesting an increase in the use of NHEJ during differentiation.

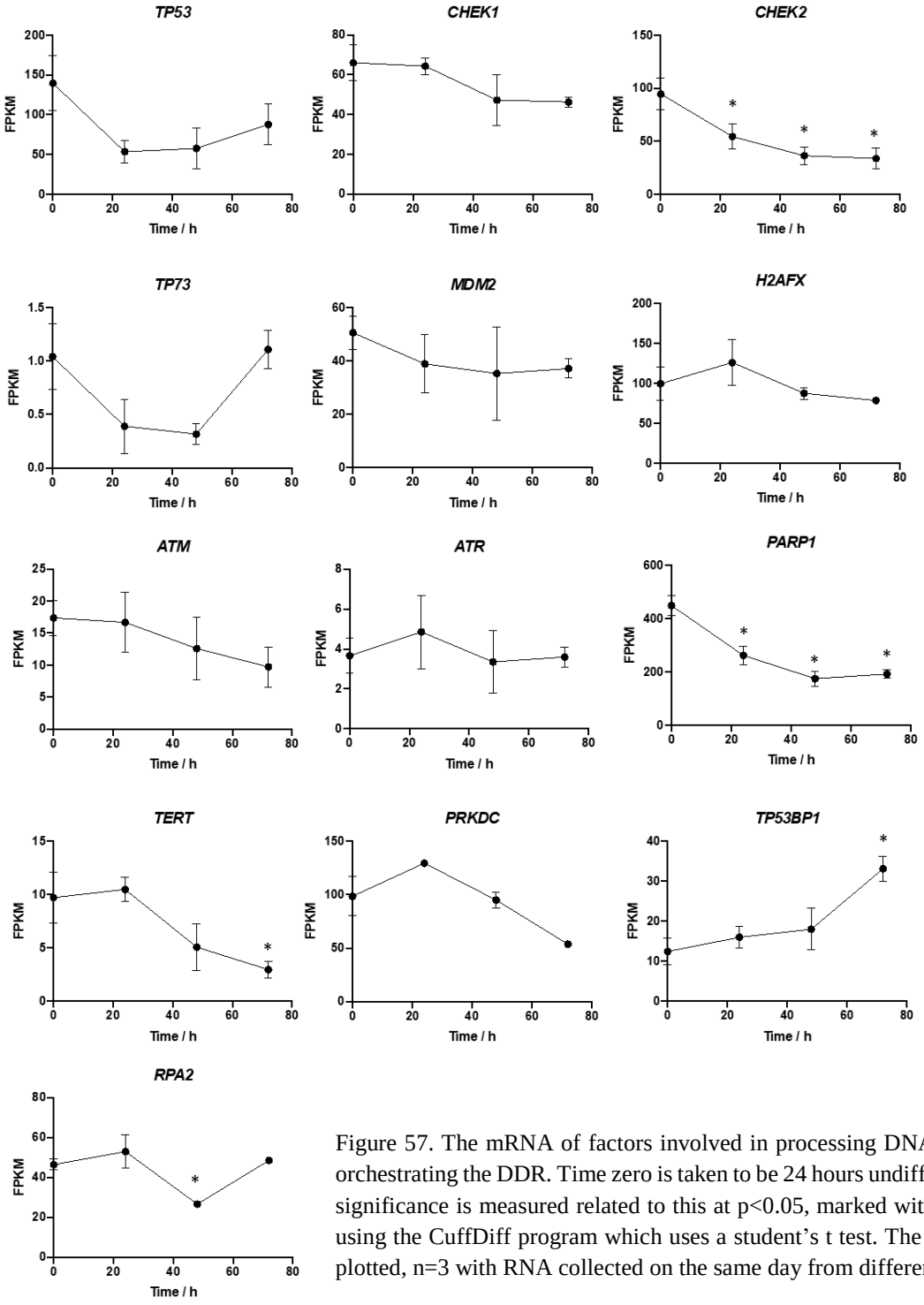


Figure 57. The mRNA of factors involved in processing DNA damage and orchestrating the DDR. Time zero is taken to be 24 hours undifferentiated and significance is measured related to this at $p<0.05$, marked with * calculated using the CuffDiff program which uses a student's t test. The mean±SD are plotted, n=3 with RNA collected on the same day from different cultures.

4.4.2 Gene expression differences between untreated wildtype and *TP53*$^{-/-}$ cells

Before analysing the changes upon treatment with MMS, the differences between the wildtype and *TP53*$^{-/-}$ cells in both the undifferentiated state and throughout differentiation were compared.

4.4.2.1 Differences in the undifferentiated state

Using the sequencing data from undifferentiated wildtype and *TP53* knockout undifferentiated cells (in Results IV) there were over 50 genes differentially expressed and this is shown in the scatterplot in Figure 58. Using these genes, the ten most significant gene ontology (GO) terms associated with the differences between these cells were analysed (Table 8). TRANSFAC TFBS (genexplain) was also used to analyse these deregulated genes for transcription factor binding site motifs (Table 9). These enrichment analyses showed that the gene expression changes were in p53-related pathways, suggesting that the genetic knockout prevented upregulation of genes with p53-target binding sites. This was expected for these undifferentiated cells, and it was clear that the lack of p53 did not alter the expression of pluripotency genes upregulated in the undifferentiated state. While there were no transcription factor binding sites with significant p-values ($p<0.05$), probably due to the small number of deregulated transcripts, the most significant results were members of the p53 family, again suggesting that these cells are true genetic knockouts.

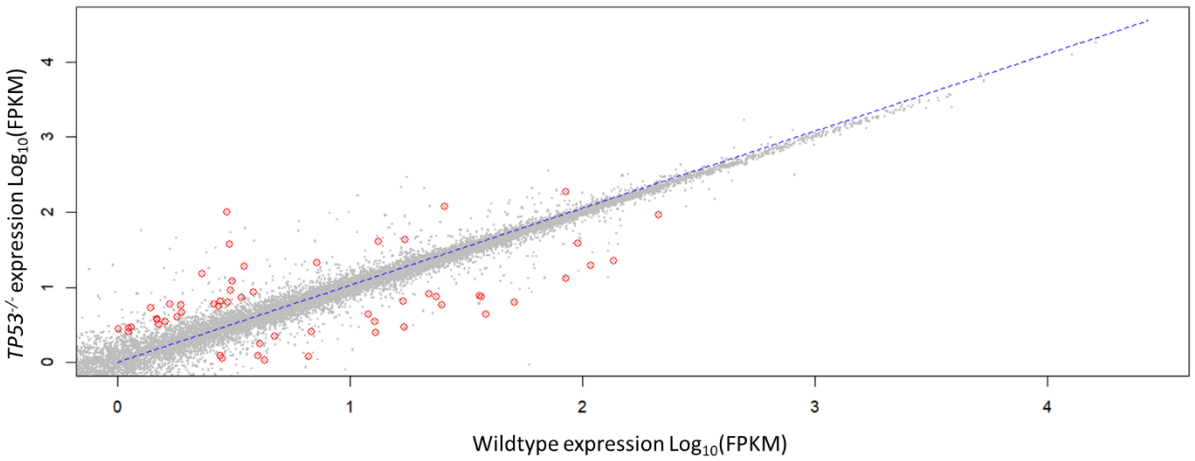


Figure 58. A scatter plot to show the genes with differential expression in the wildtype compared to the *TP53*$^{-/-}$ cell line in the undifferentiated state. Significantly differentially expressed genes are marked in red and these occur off the line x=y.

GO ID	Description	Count	Adjusted p-value
0043068	Positive regulation of programmed cell death	11/628	0.0000769
0043065	Positive regulation of apoptotic processes	11/622	0.0000769
0045569	TRAIL binding	3/5	0.0000769
0010942	Positive regulation of cell death	11/682	0.000106
0072332	Intrinsic apoptotic signaling pathway by p53 class mediator	5/82	0.000381
0006919	Activation of cysteine-type endopeptidase activity involved in apoptotic process	5/82	0.000381
0097193	Intrinsic apoptotic signaling pathway	7/295	0.00113
0070242	Thymocyte apoptotic pathway	3/17	0.00129
0009612	Response to mechanical stimulus	6/212	0.00129
0097190	Apoptotic signaling pathway	9/601	0.00129

Table 8. The ten most significant GO terms associated with the list of genes with deregulated expression between wildtype and *TP53*$^{-/-}$ cells in the undifferentiated state.

Term ID	Description	Count	Adjusted p-value
M01651	Factor: p53; motif: RGRCATGYCYRGRCATGYYY	15/2249	0.0963
M01651_0	Factor: p53; motif: RGRCATGYCYRGRCATGYYY; match class: 0	15/2249	0.0963
M01656_1	Factor: p63; motif: RRACATGTCNRGACATGTYY; match class: 1	17/2681	0.0963
M09629_1	Factor: IRF-1; motif: NAAANNGAAAGTGAAASTRN; match class: 1	3/35	0.0963

Table 9. A list of the most significant transcription factor binding sites in the genes deregulated between wildtype and *TP53*$^{-/-}$ cells.

This was then confirmed using the undifferentiated collected *TP53*$^{-/-}$ cells collected every 24 hours in the MMS endoderm experiment and compared to the wildtype cells. The Venn diagram is shown in Figure 59. There was the greatest difference at 72 hours, probably when the cells were slightly overconfluent, and 21 genes deregulated at every timepoint. The GO analysis and transcription factor binding site analysis was almost identical to that shown above in Table 8 and Table 9. The majority of differences were p53-specific, as expected.

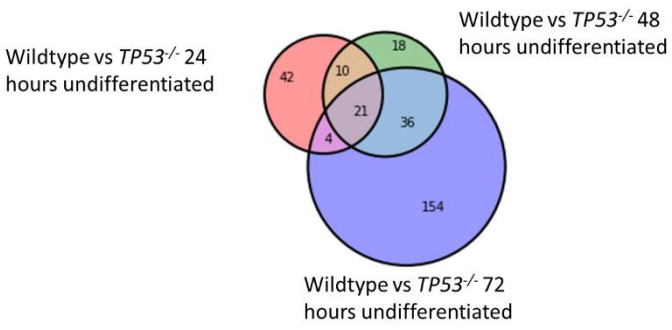
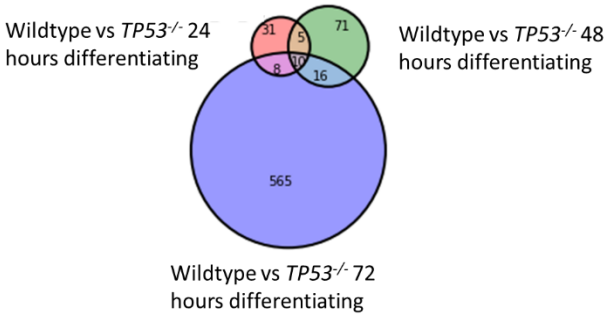


Figure 59. Comparison of the deregulated genes between wildtype and *TP53*$^{-/-}$ at 24-, 48- and 72-hours in the undifferentiated state in Venn diagram format.

4.4.2.2 Differences between wildtype and *TP53*$^{-/-}$ cells during differentiation

As in the undifferentiated state, the gene expression profiles of *TP53*$^{-/-}$ cells were compared to wildtype cells every 24 hours during differentiation (Figure 60a). As in the undifferentiated cells, the greatest change was seen at 72 hours, with only ten genes shared between the three conditions. The GO analysis overlap in these conditions is shown in Table 10, there were no significant transcription factor binding sites in this gene list. These terms were very similar to the undifferentiated state and are all directly relevant to the lack of p53. At 72 hours, 37 genes were shared between the undifferentiated and differentiated, suggesting these are specific for lack of p53 (Figure 60b). However, there were clearly many more deregulated genes during differentiation than in the undifferentiated state.

(a)



(b)

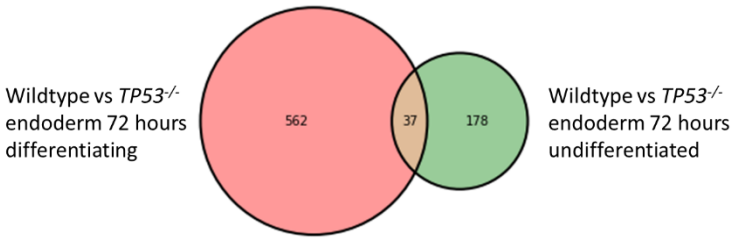


Figure 60. Venn diagrams to show the comparison of deregulated genes in the wildtype versus *TP53*$^{-/-}$ cell lines (a) every 24 hours during differentiation (b) comparing the undifferentiated and differentiated state at 72 hours.

GO term	Description	Count	Adjusted p-value
GO:0097193	Intrinsic apoptotic signaling pathway	4/295	0.00113
GO:0072332	Intrinsic apoptotic signaling pathway by p53 class mediator	3/82	0.00113
GO:0043068	Positive regulation of programmed cell death	4/628	0.00315
GO:0044065	Positive regulation of apoptotic process	4/622	0.00315
GO:0010952	Positive regulation of peptidase activity	3/179	0.00315
GO:2001056	Positive regulation of cysteine-type endopeptidase activity	3/162	0.00315
GO:0042772	DNA damage response, signal transduction resulting in transcription	3/135	0.00315
GO:006978	DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator	2/20	0.00315
GO:0097190	Apoptotic signaling pathway	2/19	0.00315
GO:0010950	Positive regulation of endopeptidase activity	4/601	0.00315

Table 10. The GO enrichment of the ten transcripts deregulated between the wildtype and *TP53*$^{-/-}$ cells at 24-, 48- and 72-hours in differentiation.

Comparing the wildtype and *TP53*$^{-/-}$ cells at 72 hours into the differentiation had over 500 genes and the GO enrichment and transcription factor analysis are shown in Table 11 and Table 12. More of the gene ontology terms were enriched in developmental processes, suggesting that *TP53*$^{-/-}$ cells may have slightly altered differentiation. This may be due to the confluence of these cells, or the lack of cell death experienced during normal differentiation in these cells. There were a number of enriched transcription factor binding sites, however, none of them were in the p53 family of transcription factors. This suggested that *TP53*$^{-/-}$ was not the key transcription factor regulating these changes.

GO term	Description	Count	Adjusted p-value
GO:0048856	Anatomical structure development	220/5793	8.29×10^{-8}
GO:0007275	Multicellular organism development	206/5321	8.29×10^{-8}
GO:0032502	Developmental process	230/6212	1.36×10^{-7}
GO:0009653	Anatomical structure morphogenesis	120/2598	1.66×10^{-7}
GO:1905114	Cell surface receptor signaling pathway involved in cell-cell signaling	43/569	5.57×10^{-7}
GO:0016477	Cell migration	78/1466	0.00000102
GO:0009888	Tissue development	94/1926	0.00000107
GO:0023051	Tube morphogenesis	56/899	0.00000107
GO:0023051	Regulation of signaling	144/3482	0.00000141
GO:0048646	Anatomical structure formation involved in morphogenesis	65/1144	0.00000143

Table 11. GO enrichment of significantly differentially expressed genes between wildtype and *TP53*$^{-/-}$ cells at 72 hours during differentiation.

TRANSFAC TFBS	Factor	Motif	Count	Adjusted p-value
TF:M07039	ETF	CCCCGCCCCYN	418/13652	$5.38x10^{-7}$
TF:M00986	Churchill	CGGGNN	432/14245	$5.38x10^{-7}$
TF:M03876	Kaiso	GCMGGGTGCRGS	307/9094	$5.38x10^{-7}$
TF:M07040	GKLF	NNRRGRRNGNSNNN	397/12747	$5.38x10^{-7}$
TF:M09723	BTEB1	GGGGGCGGGGCNGSGGNGS	325/9991	0.0000048

Table 12. Transcription factor binding site enrichment analysis of genes deregulated between wildtype and *TP53*$^{-/-}$ cells at 72 hours in differentiation.

4.4.3 Treatment with MMS in wildtype and *TP53*$^{-/-}$ cells

The differences in the wildtype and *TP53*$^{-/-}$ cells after treatment with MMS was compared in both the undifferentiated state and during differentiation.

4.4.3.1 Treatment with MMS in the undifferentiated state

MMS was added to the undifferentiated and differentiating cells at 24 hours and then the cells were collected at 48 and 72 hours. The changes in gene expression in wildtype and *TP53*$^{-/-}$ cells in the undifferentiated state are shown in Figure 61a and b. Very few genes were deregulated at 48 hours and 72 hours in the undifferentiated state of both cell lines when the cells were treated with MMS. This explained the low level of death and lack of any observable changes in these cells. Interestingly, there were fewer changes at 72 hours, suggesting that the cells have processed any damage by this point, this is on the contrary to the differentiating cells where the differences were greater at 72 hours. There were also only five deregulated genes which overlapped between the wildtype MMS treated and *TP53*$^{-/-}$ MMS treated cells at 48 hours, suggesting common pathways are not triggered possibly due to the lack of p53.

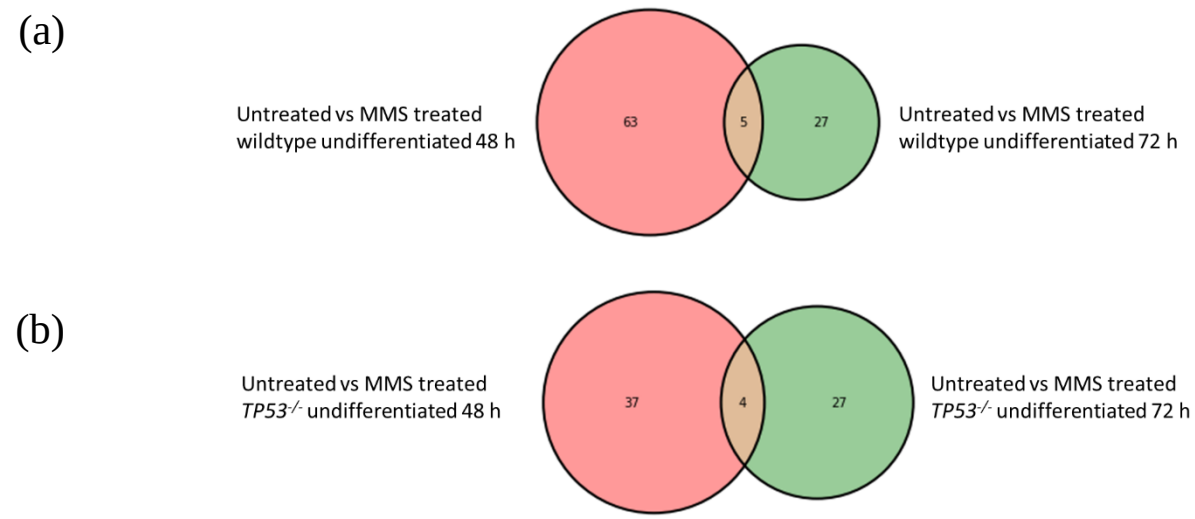


Figure 61. Venn diagram comparisons of changes in the undifferentiated state after MMS treatment (a) comparison of 48 hours to 72 hours in the wildtype cell line (b) as (a) but with *TP53*$^{-/-}$. Venn diagrams not to scale between cell lines.

The enrichment in the MMS treated cells compared to the untreated cells at 48 hours were analysed for GO terms in both *TP53*$^{-/-}$ and wildtype (Table 13). There were no significantly enriched transcription factor binding sites in these gene lists. In the wildtype cell line these GO terms were associated with responses to external stimuli and cell death, as well as development. Suggesting that MMS treatment induces a number of significant responses in these cells. The enrichment showed different pathways deregulated in *TP53*$^{-/-}$ cells, probably due to the lack of p53. These terms were associated with calcium signalling pathways and muscle hypertrophy, which p53 has been previously implemented in (Mak et al., 2017).

GO term	Description	Count	Adjusted p-value
Wildtype			
GO:0048513	Animal organ development	26/3428	0.000745
GO:0009605	Response to external stimulus	24/3141	0.000794
GO:1901700	Response to oxygen-containing compound	35/6513	0.000794
GO:0007154	Cell communication	16/1571	0.000794
GO:0065009	Regulation of molecular function	20/2332	0.000794
GO:0023052	Signaling	36/6554	0.000794
GO:0009893	Positive regulation of metabolic process	24/3333	0.000794
GO:0010604	Positive regulation of macromolecule metabolic process	23/3134	0.000794
GO:0010941	Regulation of cell death	15/1378	0.000794
GO:0043067	Regulation of programmed cell death	17/1676	0.000794
TP53$^{-/-}$			
GO:0032501	Multicellular organismal process	24/7414	0.00794
GO:0014896	Muscular hypertrophy	4/109	0.00794
GO:0014897	Striated muscle hypertrophy	4/107	0.00794
GO:0003300	Cardiac muscle hypertrophy	4/105	0.00794
GO:0007267	Cell-cell signaling	11/1559	0.00794
GO:0099094	Ligand-gated cation channel activity	4/103	0.00794
GO:0005219	Ryanodine-sensitive calcium-release channel activity	2/4	0.00794
GO:0048763	Calcium-induced calcium release activity	2/4	0.00794
GO:0060079	Excitatory postsynaptic potential	4/113	0.00814
GO:0071495	Cellular response to endogenous signal	10/1347	0.00815

Table 13. The GO enrichment analysis of the genes deregulated in the MMS treated cells compared to the untreated cells at 48 hours in both undifferentiated wildtype and *TP53*$^{-/-}$ cell lines. The ten most significant terms are shown.

The differences in the MMS treated *TP53*$^{-/-}$ treated cells compared to the MMS treated wildtype cells at both 48 and 72 hours had a high level of overlap (Figure 62). This suggested that there were distinct differences in the transcriptome in wildtype versus *TP53*$^{-/-}$ cells when treated with damaging agents. These differences had significant enrichment in stress, checkpoints and p53 pathways, shown in Table 14. The enriched transcription factor binding sites are shown in Table 15. Clearly the lack of p53 was causing big differences in the control of DNA damage signalling pathways, and p53 was one of the key transcription factors implicated in these differences.

GO term	Description	Count	Adjusted p-value
GO:0072331	Signal transduction by p53 class mediator	14/223	$3.39x10^{-8}$
GO:0072332	Intrinsic apoptotic signaling pathway by p53 class mediator	10/82	$3.39x10^{-8}$
GO:0009628	Response to abiotic stimulus	25/1120	$8.34x10^{-7}$
GO:0044783	G1 DNA damage checkpoint	8/66	0.0000135
GO:0044819	Mitotic G1/S transition checkpoint	8/65	0.0000135
GO:0031571	Mitotic G1 DNA damage checkpoint	8/65	0.0000135
GO:0033554	Cellular response to stress	31/1870	0.00000272
GO:0097193	Intrinsic apoptotic signaling pathway	13/295	0.00000272
GO:0034644	Cellular response to UV	8/77	0.00000313
GO:0006974	Cellular response to DNA damage stimulus	20/815	0.00000355

Table 14. The GO enrichment of genes deregulated at both 48 and 72 hours in the undifferentiated state with MMS treatment in *TP53*$^{-/-}$ compared to wildtype cells.

TRANSFAC TFBS	Factor	Motif	Count	Adjusted p-value
TF:M01873	Egr-1	GCGGGGGCGG	62/6728	0.00119
TF:M01655	p53	GGACATGYYCGGACATGYYC	40/3503	0.00173
TF:M02089	E2F-3	GGCGGGN	93/13182	0.00235
TF:M08878	EGF	CGCCCCCGCNN	64/7482	0.00286
TF:M07354	Egr-1	GCGGGGGCGG	65/7783	0.0043

Table 15. Transcription factor binding site enrichment of genes overlapping in Figure 62.

Figure 62. A Venn diagram to show the comparison of the deregulated genes in the undifferentiated state in the wildtype and *TP53*$^{-/-}$ cell lines after MMS treatment at 48 and 72 hours.

4.4.3.2 MMS treatment during differentiation

Comparing the untreated and MMS treated differentiating samples at 48 and 72 hours in the wildtype and *TP53*$^{-/-}$ cell lines showed that there were many more deregulated genes at both times in the wildtype cell line (Figure 63). There were also more changes at 72 hours compared to 48 hours in the wildtype cell line suggesting that the treatment continues to exacerbate the phenotype over 24 hours after treatment. In the *TP53*$^{-/-}$ cells there were many fewer deregulated genes at both times, and a very small overlap between the times. There were also fewer genes deregulated at 72 hours compared to 48 hours in this cell line, again suggesting that differentiation is not affected in the same manner. The GO analysis for the difference between the treated and untreated, differentiating cells at 72 hours were analysed in each cell line (Table 16) and the significant transcription factor binding sites for these are shown in Table 17, there were no significant sites in the *TP53*$^{-/-}$ gene list. The GO terms in the wildtype deregulated transcripts were mainly associated with development and morphogenesis whereas in *TP53*$^{-/-}$ they were mainly transporters and channels. Comparing the wildtype and *TP53*$^{-/-}$ deregulated genes at both 48 and 72 hours there were low levels of overlap, suggesting that some of the differences were the same with or without MMS (Figure 64).

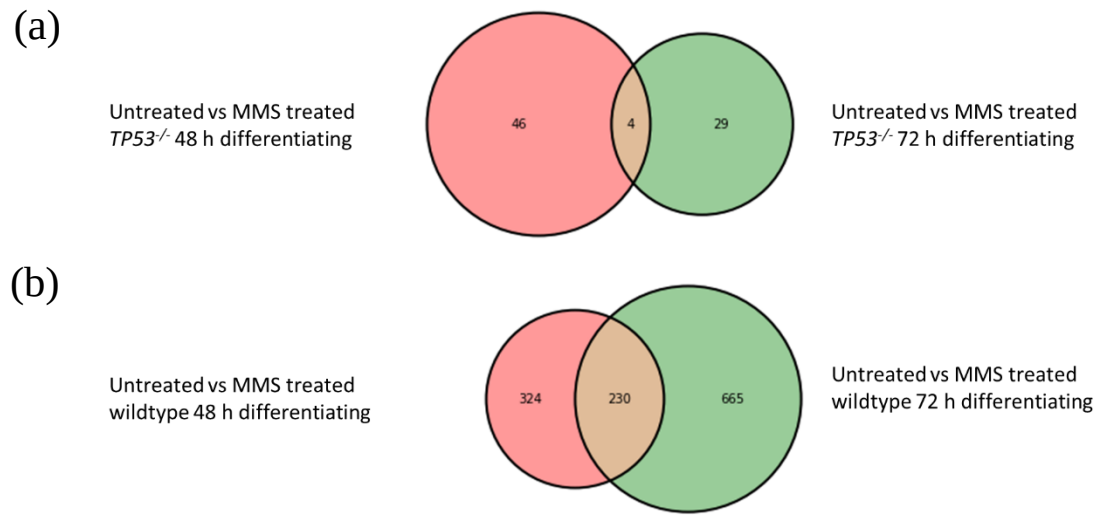


Figure 63. Venn diagrams to show the comparison of deregulated genes after MMS treatment during differentiation at 48 hours and 72 hours (a) *TP53*$^{-/-}$ (b) wildtype. Venn diagram scales are not comparative between cell lines.

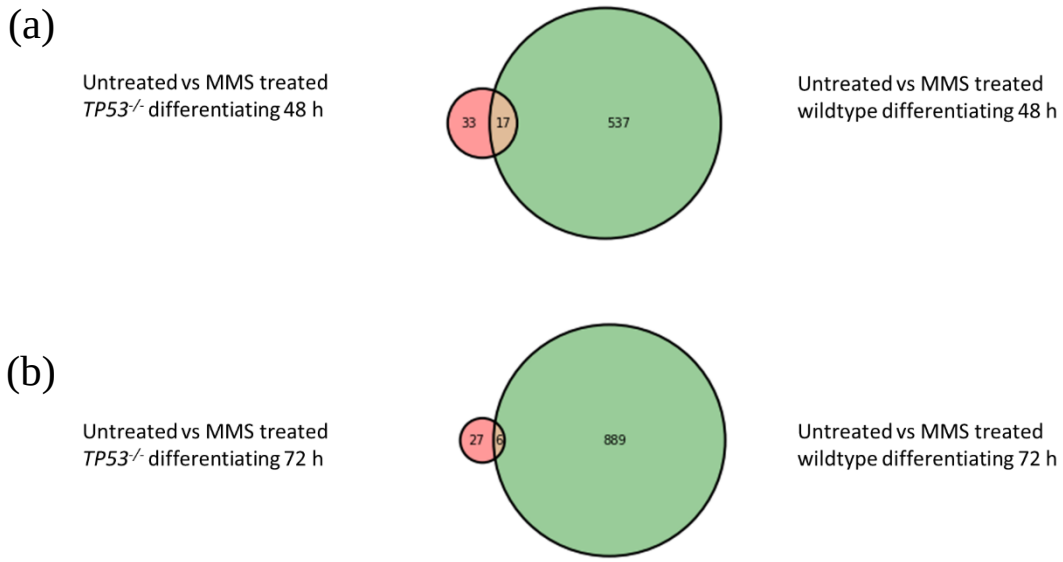


Figure 64. Venn diagram comparison of wildtype and *TP53*$^{-/-}$ MMS treated cells at (a) 48 hours and (b) 72 hours.

GO term	Description	Count	Adjusted p-value
Wildtype 72 h			
GO:0009653	Anatomical structure morphogenesis	241/2598	$1.75x10^{-34}$
GO:0048856	Anatomical structure development	400/5793	$1.39x10^{-33}$
GO:0007275	Multicellular organism development	378/5321	$1.39x10^{-33}$
GO:0032502	Developmental process	413/6212	$2.19x10^{-31}$
GO:0009887	Animal organ morphogenesis	129/969	$3.71x10^{-31}$
GO:0048731	System development	336/4760	$9.98x10^{-28}$
GO:0048731	Animal organ development	265/3428	$6.23x10^{-26}$
GO:0032501	Multicellular organismal process	449/7414	$1.17x10^{-25}$
GO:0048646	Anatomical structure formation involved in morphogenesis	129/1144	$3.23x10^{-24}$
GO:0035295	Tube morphogenesis	124/1079	$6.04x10^{-24}$
TP53$^{-/-}$ 72 h			
GO:0008324	Cation transmembrane transporter activity	7/664	0.00942
GO:0046873	Metal ion transmembrane transporter activity	6/457	0.00942
GO:0022803	Passive transmembrane transporter activity	6/470	0.00942
GO:0015267	Channel activity	6/469	0.00942
GO:0005261	Cation channel activity	5/320	0.0148
GO:0005215	Transporter activity	8/1251	0.0223
GO:0015075	Ion transmembrane transporter activity	7/897	0.0223
GO:0022890	Inorganic cation transmembrane transporter activity	6/605	0.0223
GO:0015077	Monovalent inorganic cation transmembrane transporter activity	5/392	0.0223
GO:0022842	Narrow pore activity	2/19	0.0288

Table 16. GO enrichment for transcripts deregulated between untreated and MMS treatment at 72 hours of differentiation in wildtype and *TP53*$^{-/-}$ cells. MMS treatment was performed at 24 hours.

TRANSFAC TFBS	Factor	Motif	Count	Adjusted p-value
TF:M07354	Egr-1	GCGGGGGCGG	443/7783	$5.63x10^{-18}$
TF:M01104	MOVO-B	GNGGGGG	345/5567	$1.26x10^{-17}$
TF:M03876	Kaiso	GCMGGGRGCRGS	493/9094	$3.22x10^{-17}$
TF:M07040	GKLF	NNRRRGRRNGNSNNN	628/12747	$7.4x10^{-17}$
TF:M00695	ETF	GVGGMGG	404/7004	$7.4x10^{-17}$

Table 17. Significantly enriched transcription factor binding sites in genes deregulated at 72 hours in wildtype MMS treated cells compared to wildtype untreated cells during differentiation. There were no significant results in the *TP53*$^{-/-}$ cells.

The genes differentially expressed between the wildtype and *TP53*$^{-/-}$ differentiating cells after MMS treatment were compared at 48 hours and 72 hours (Figure 65). This showed that there was a large overlap between these times, suggesting the differences in how the two cell lines deal with damage. The GO term enrichment for this overlap is shown in Table 18. These terms were mainly responses due to oxygen levels, possibly due to the different ways of processing DNA damage, and developmental pathways. The transcription factor binding site enrichment is shown in Table 19.

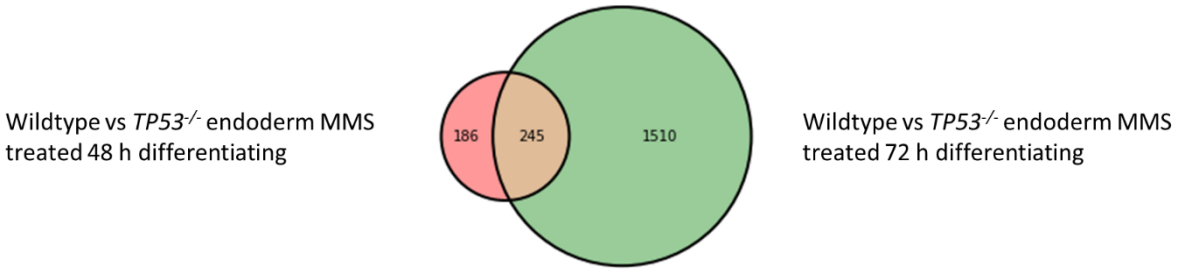


Figure 65. Venn diagram to compare the differences at 48- and 72-hours between the wildtype and *TP53*$^{-/-}$ cell lines after treatment with MMS during differentiation.

GO term	Description	Count	Adjusted p-value
GO:0070482	Response to oxygen levels	23/339	$1.5x10^{-8}$
GO:0036293	Response to decreased oxygen levels	22/316	$1.5x10^{-8}$
GO:0009628	Response to abiotic stimulus	40/1120	$3.34x10^{-8}$
GO:0001666	Response to hypoxia	21/306	$3.34x10^{-8}$
GO:0007275	Multicellular organism development	101/5321	$2.51x10^{-7}$
GO:0048856	Anatomical structure development	106/5793	$4.64x10^{-7}$
GO:0031012	Extracellular matrix	21/376	$7.74x10^{-7}$
GO:0009653	Anatomical structure morphogenesis	61/2598	0.00000177
GO:0032502	Developmental process	109/6212	0.00000187
GO:0050840	Extracellular matrix binding	9/53	0.00000278

Table 18. GO term enrichment of the overlap at 48 and 72 hours of deregulated genes in the *TP53*$^{-/-}$ MMS treated cells compared to wildtype treated cells during differentiation. Cells were treated with MMS at 24 hours.

TRANSFAC TFBS	Factor	Motif	Count	Adjusted p-value
TF:M01219	SP1:SP3	CCGCCCCCYCC	122/6823	$2.86x10^{-7}$
TF:M00986	Churchill	CGGGNN	195/14245	0.00000133
TF:M00333	ZF5	NRNGNGCGCGCWN	201/15142	0.00000338
TF:M02089	E2F-3	GGCGGGN	184/13182	0.00000338
TF:M00982	KROX	CCCGCCCCCRCCCC	129/7814	0.00000338

Table 19. Transcription factor binding site enrichment of the overlap at 48 and 72 hours of deregulated genes in the *TP53*$^{-/-}$ MMS treated cells compared to wildtype treated cells during differentiation.

4.5 Preliminary data using human ES cells show a similar trend on treatment with MMS

4.5.1 Endoderm differentiation

The BOBSC cells are hiPS cells, not hES cells, and are known to have a single translocation near the *MYC* gene. To understand whether the observations from these cells can be generalised to a greater extent across differentiation the key experiments were repeated in the H9 hES cell line. Ideally these would also be repeated in the BOBSC cell line without the translocation. The cells were initially differentiated without any treatment to analyse their differentiation capability (Figure 66a). This showed that they differentiate with a high efficiency under the same conditions as the hiPS cells, but the level of EOMES at 72 hours remained higher in the H9 cells. Further experiments were compared.

4.5.2 MMS treatment

Initial observations suggested that the undifferentiated H9 cells did not tolerate MMS treatment as well as the BOBSC cells: there was a greater level of cell death using 5 ppm MMS. However, the endoderm differentiated cells were still much sicker than the undifferentiated cells, as in the hiPSCs. H9s were differentiated with MMS treatment, either one dose at 24 hours or three single doses at 0-, 24- and 48-hours (Figure 66b). When the cells were administered three doses, the H9 cells were much sicker than the BOBSC cells in both the undifferentiated and differentiated state, possibly suggesting that the dose many need to be decreased to see the same effect. There was also notably less G2/M arrest in the 0-, 24- and 48-hour MMS treated H9 cells but this may have been biased due to the increased cell death. However, MMS clearly decreased the expression of SOX17 during the differentiation in the H9 cells, although possibly to a lesser extent. Suggesting that DNA damage prevents differentiation in H9s and may be a more general phenomenon.

4.5.3 The spike of H2A.X phosphorylation in differentiation

The cells were also analysed between 48 and 52 hours to see if there was a spike of γH2A.X around the time that this was seen in the BOBSC cells (Figure 66c). This shows that there was an upregulation of phospho-H2A.X at this time, suggesting that this is specific to the differentiation and showing that the H9s act in a similar manner to the BOBSC cells. However, the level of upregulation appeared to be lower in the H9s compared to the BOBSC cell line. This is likely to be a general phenomenon of cells moving through the EMT.

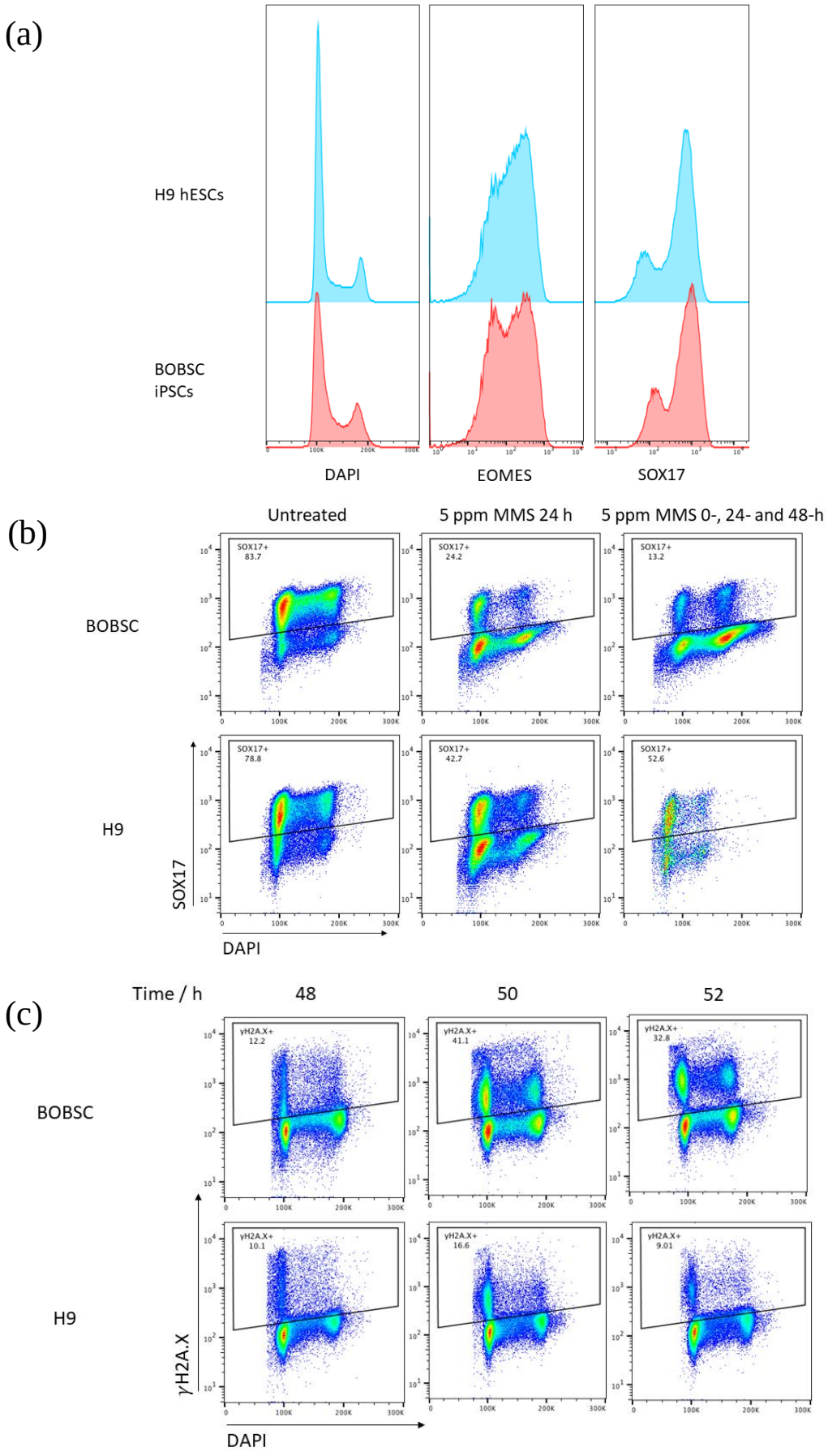


Figure 66. (See next page.) Analysis of the H9 hESCs in comparison to the BOBSC hiPSC cell line (a) unperturbed differentiation monitored by DAPI, EOMES and SOX17 at 72 hours (b) SOX17 expression at 72 hours after MMS treatment as labelled (c) phosphorylation of H2A.X at the point of the spike during differentiation.

Chapter 5. Results III: The effect of G-quadruplex-binding ligands in the undifferentiated state and during differentiation of hiPSCs

5.1 Culturing hiPSCs with G4 ligands in the pluripotent state

5.1.1 Maintaining the cells in the undifferentiated state

Having analysed the effect of inducing DNA damage on the endoderm differentiation capacity of hiPSCs, I wanted to understand how growing and differentiating the cells with G4 ligands would compare to the DNA damaging agent treatment. BOBSC cells were cultured in the undifferentiated state with the G4 ligands PDS and PhenDC3 (De Cian et al., 2007; Rodriguez et al., 2008), discussed previously in Results I, to see how well the cells were able to tolerate the ligands. Concentrations up to 4 μM of both ligands were added to cells in the undifferentiated state for five days and there were minimal differences in doubling times suggesting that these concentrations could be used further for differentiation experiments. However, as in the mESCs, there were more 'floating cells' in the wells treated with G4 ligands, possibly also suggesting a loss of pluripotency or increased cell death. The following experiments were performed using PDS or PhenDC3 at a concentration where the DMSO was known not to alter the outcome of differentiation and had previously been shown to stabilise G4s in DT40 cells (Guilbaud et al., 2017).

5.1.2 The cell cycle profile

To understand whether the G4 ligands were causing replication stress, as has been previously shown (Guilbaud et al., 2017), undifferentiated BOBSC cells were grown with either PDS or PhenDC3 for two days and EdU labelling was performed, as previously. The proportion of cells in each phase of the cell cycle with and without the ligand was quantified (Figure 67), replicates are shown in comparison to differentiation (Figure 69b). Culturing the cells in PDS in the undifferentiated state caused a very slight increase in the proportion of cells in G2/M, as would be expected, but did not suggest that there was a block in the cell cycle that would alter doubling time significantly. This result is in agreement with previous data showing that PDS can induce DNA damage (Rodriguez et al., 2012). The PhenDC3 treated cells showed no increase in G2/M

arrest compared to both the untreated and DMSO treated conditions. This suggested that PhenDC3 was not causing a replication impediment in a similar manner to PDS.

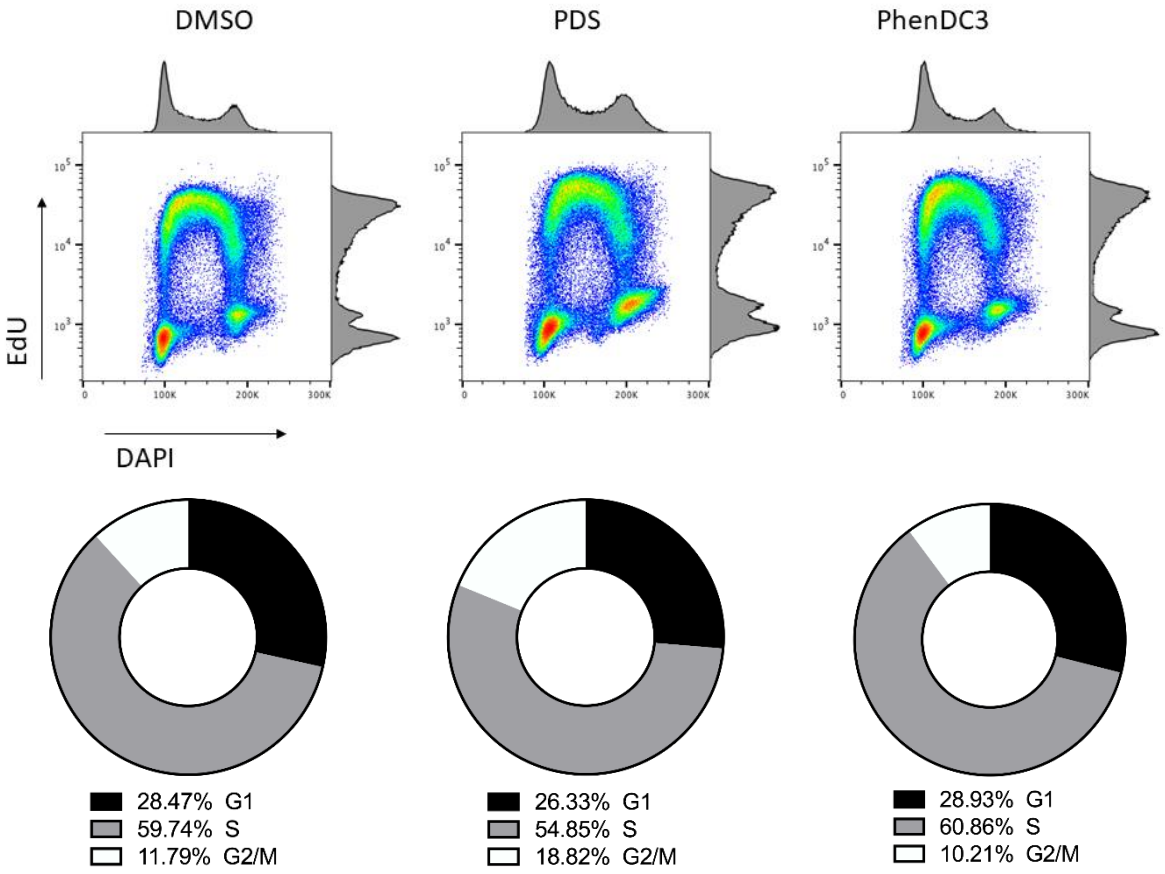


Figure 67. Cells were cultured in the undifferentiated state for 50 hours with 1 µM PDS, PhenDC3 or an equivalent volume of DMSO. EdU was added for incorporation into the synthesising DNA for one hour prior to cell collection. Cells were then collected and fixed for flow cytometry and the Click-iT reaction was performed (above). The number of cells in each phase was quantified and plotted for each condition (below), G1 (black), S (grey) and G2/M (white).

5.1.3 The DNA damage response to G4 ligands

To see whether the G4 ligands were increasing DNA damage in treated cells, I analysed the level of γH2A.X using permeabilised flow cytometry. Cells were grown in the undifferentiated state with PDS or PhenDC3 for 50 hours, as above, and collected for flow cytometry analysis of γH2A.X (Figure 68a). The data show that culturing in PDS did cause a shift in the level of phospho-H2A.X in cells compared to the mock-treated and PhenDC3 treated cells. However, the level of increase of γH2A.X in PDS treated cells was much smaller than of those cells treated with HU, suggesting that the level of damage was minimal. Interestingly, PhenDC3 did not cause an increase in this DNA damage marker, further suggesting that PDS and PhenDC3 were not acting in an exactly analogous manner in the cells.

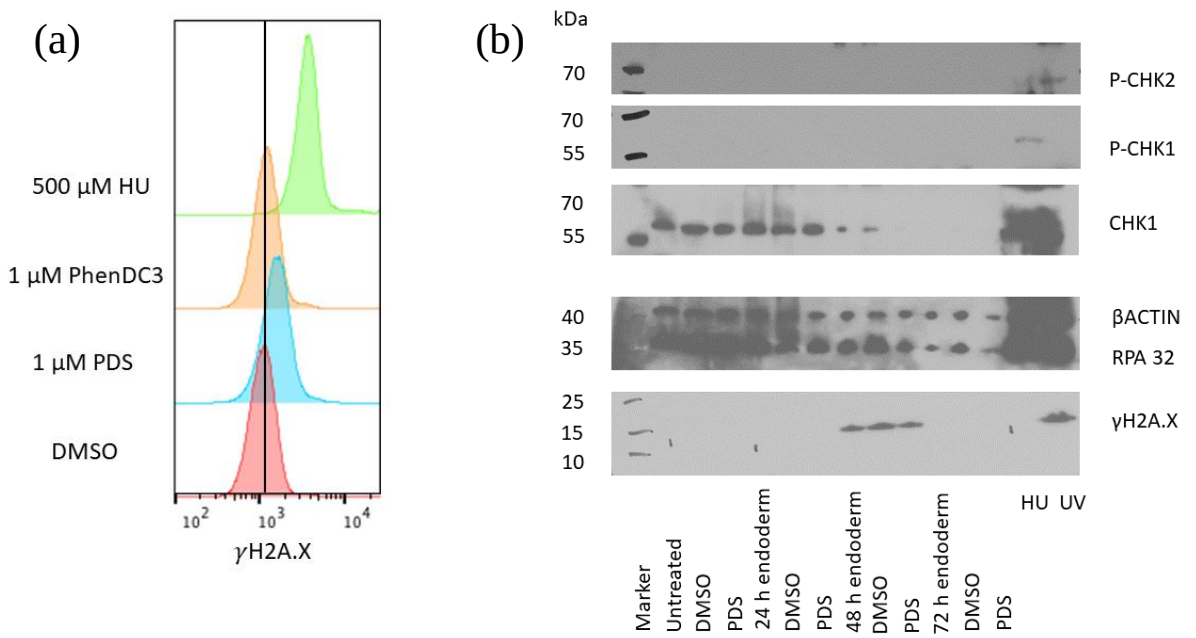


Figure 68. Analysis of phosphorylation of DDR proteins after treatment with G4 ligands (a) flow cytometry analysis of the phosphorylation of H2A.X in the BOBSC cells in the undifferentiated state after treatment with 1 µM PDS or PhenDC3 for 50 hours, to compare with the spike of phosphorylation during differentiation. HU treatment was included as a positive control. The median level with mock-treated cells is shown with a line. (b) western blotting analysis of the phosphorylation of CHK1, CHK2 and H2A.X in 2 µM PDS treated cells. HU and UV treatment were used as positive controls.

Following γH2A.X, levels of pCHK2 and pCHK1, were analysed using western blotting to check whether there was any upregulation of the DDR in PDS treated cells (Figure 68b). Only the positive control cells treated with 2 mM HU gave a pCHK2 signal on the blot and the UV treated sample gave a positive pCHK1 signal on the blot. This suggested that PDS was not causing phosphorylation of checkpoint proteins at a level detectable by western blotting during differentiation. Phosphorylation of H2A.X could only be detected by flow cytometry, showing that the level of phosphorylation was low. Despite both ligands being shown to induce G4-dependent replication stress (Guilbaud et al., 2017), only PDS triggers H2A.X phosphorylation and cell cycle perturbation in this system.

5.1.4 RNA expression changes in the pluripotent state induced by G4 ligands

BOBSC cells were cultured in the undifferentiated state with 2 µM PDS, PhenDC3 or the equivalent volume of DMSO for three days. RNA was extracted every 24 hours to understand whether gene expression changes were induced by G4 ligands in the undifferentiated state. The RNA was converted into DNA libraries and sequenced. The ligands induced a change in gene expression and this data will be discussed in Section 5.3 in comparison to the changes induced in differentiation.

5.2 G4 ligand treatment during definitive endoderm differentiation

5.2.1 Changes induced by G4 ligands during endoderm differentiation

5.2.1.1 The cell cycle profile

Throughout definitive endoderm differentiation the cell cycle profile changes from cells being mostly in S phase of the cell cycle to mainly in G1, as is indicative of a more differentiated cell type. When cells were differentiated in the presence of PDS there was a clear hold up of cells in G2/M phase at 24 hours, shown in the DAPI plot in Figure 69, and at 72 hours a smaller proportion of cells were in G1 phase. While growing the undifferentiated cells in PDS (Figure 67) may have slightly increased the proportion of cells in G2/M at any one time, this effect was much more pronounced during differentiation (Figure 69b). However, when PhenDC3 was used during differentiation there was a minimal change in the cell cycle profile until 72 hours where there were a higher proportion of cells in G1 phase typical of a differentiated cell (Figure 69a and c), this is discussed in more detail in Section 5.2.2.2. This further suggested that the ligands could be acting in different manners during differentiation.

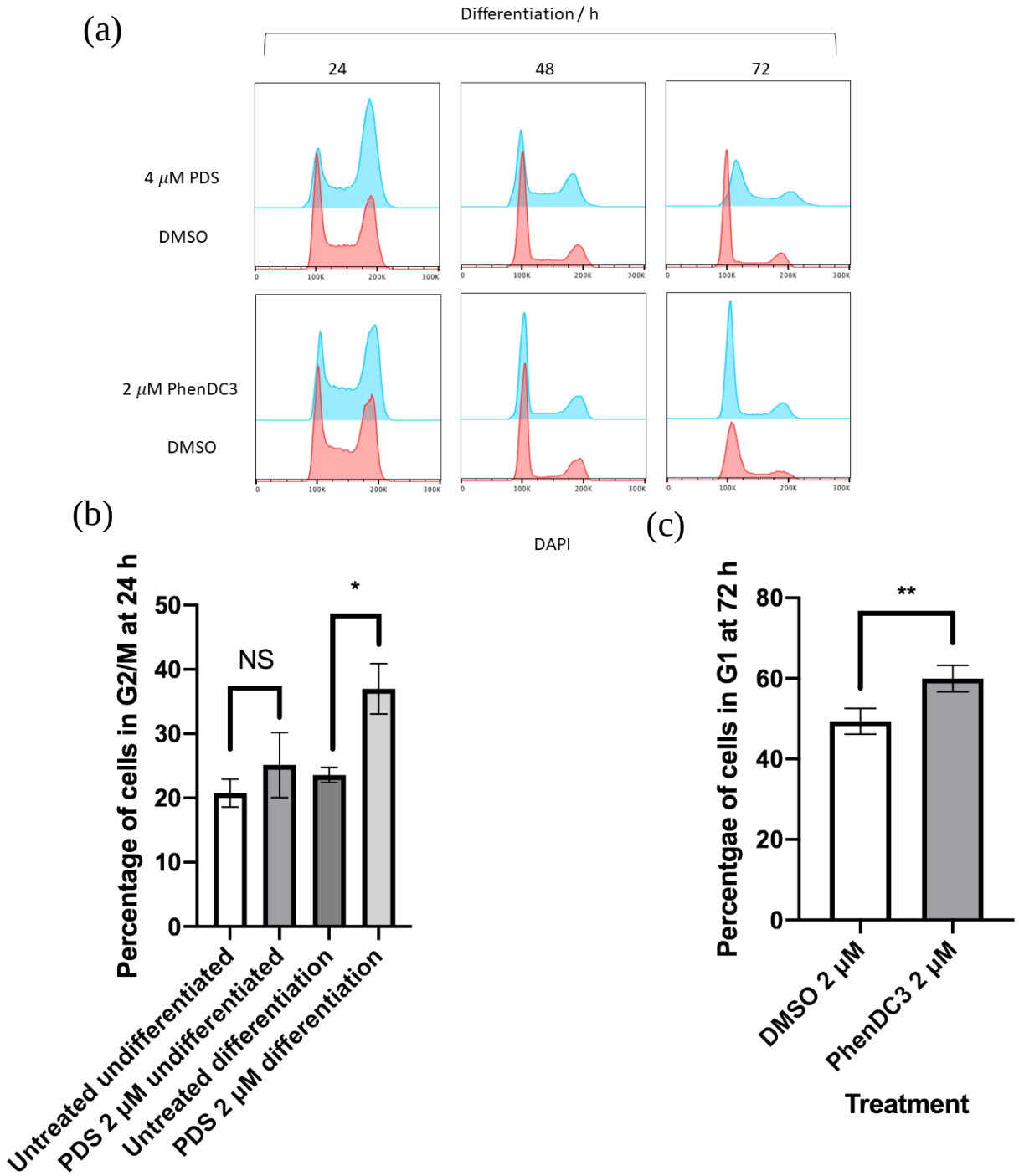
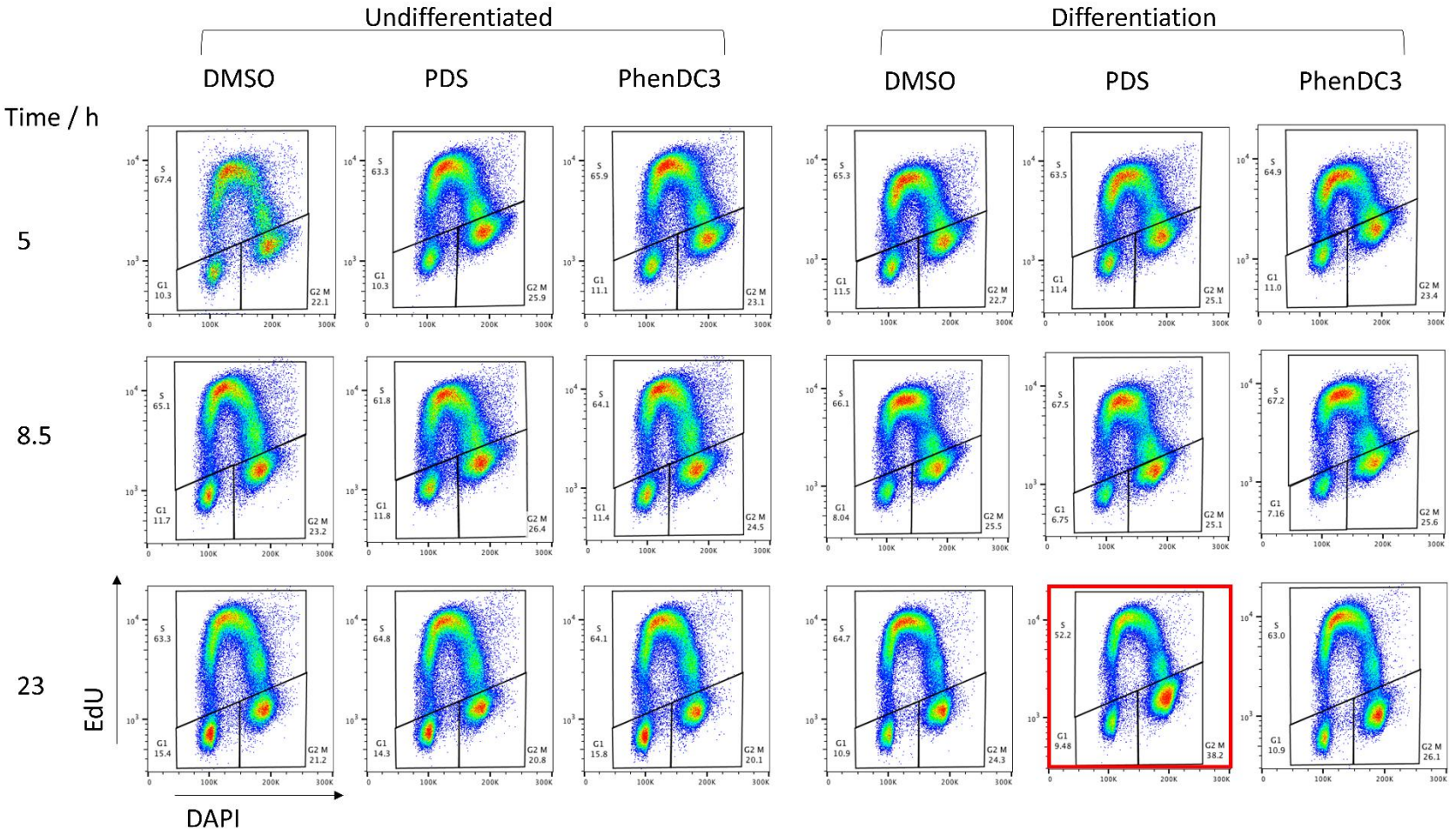


Figure 69. Cell cycle changes during differentiation in the presence of PDS and PhenDC3 G4 ligands monitored using DAPI staining to analyse the DNA content. (a) Ligand treated samples are shown in blue and mock-treated in red (b) replicate values of the percentage of cells in G2/M with PDS treatment either in the undifferentiated state. Mean±SEM are plotted, n=4, p=0.0172 during differentiation and not significant for the undifferentiated state using a paired t test. (c) replicate experiments to show the percentage of cells in G1 at 72 hours of differentiation with DMSO or PhenDC3 treatment, mean±SEM are plotted, n=8, p=0.0049 using a paired t test.

As the effect of PDS was so pronounced at 24 hours, earlier timepoints during differentiation were analysed using EdU incorporation, as before, and compared to DMSO and PhenDC3 treated cells. Cells were treated with ligand in the undifferentiated and differentiating states and were incubated with EdU before being collected between five and 23 hours after the initiation of differentiation (Figure 70). There was no increase in the G2/M population in PDS treatment until 23 hours, and this was not shared with the PhenDC3 treatment or, importantly, the undifferentiated cells. This further highlighted the idea, proposed in Results II, that on the transition from undifferentiated to differentiating the cells were less able to tolerate damage, if DNA damage was the means through which PDS was acting. The height of the S phase population was also slightly lower in the treated cells (Hwang et al., 2016), suggesting that cells differentiated with PDS may be synthesising DNA slower. Interestingly, there was a clear change in the speed DNA synthesis in all conditions during differentiation: the EdU fluorescence change was the smallest between G1 and S phase at five hours. This early timepoint is likely to be a key point for cell fate commitment.

Figure 70. (See next page) 2D-cell cycle analysis using EdU to analyse the proportion of cells in each phase of the cell cycle in the undifferentiated state (left) and during differentiation (right) cultured with 2 µM PDS, 2 µM PhenDC3 or an equivalent volume of DMSO. The red box shows the increase in G2/M phase of the cell cycle after PDS treatment during differentiation. Quantification is shown in the table below.

	Undifferentiated			Differentiated		
% cells	G1	S	G2/M	G1	S	G2/M
	5 h					
DMSO	10.3	67.4	22.1	11.5	65.3	22.7
2 µM PDS	10.3	63.3	25.9	11.4	63.5	25.1
2 µM PhenDC3	11.1	65.9	23.1	11.0	64.9	23.4
	8.5 h					
DMSO	11.7	65.1	23.2	8.04	66.1	23.5
2 µM PDS	11.8	61.8	26.4	6.75	67.5	25.1
2 µM PhenDC3	11.4	64.1	24.5	7.16	67.2	25.6
	23 h					
DMSO	15.4	63.3	21.2	10.9	64.7	24.3
2 µM PDS	14.3	64.8	20.8	9.48	52.2	38.2
2 µM PhenDC3	15.8	64.1	20.1	10.9	63.0	26.1


Undifferentiated
Differentiation
DMSO
PDS
PhenDC3
DMSO
PDS
PhenDC3
Time / h
5
8.5
23
EdU
DAPI
S 67.4
G1 10.3
G2 M 22.1
S 63.3
G1 10.3
G2 M 25.9
S 65.9
G1 11.1
G2 M 23.1
S 65.3
G1 11.5
G2 M 22.7
S 63.5
G1 11.4
G2 M 25.1
S 64.9
G1 11.0
G2 M 23.4
S 65.1
G1 11.7
G2 M 23.2
S 61.8
G1 11.8
G2 M 26.4
S 64.1
G1 11.4
G2 M 24.5
S 66.1
G1 8.04
G2 M 25.5
S 67.5
G1 6.75
G2 M 25.1
S 67.2
G1 7.16
G2 M 25.6
S 63.3
G1 15.4
G2 M 21.2
S 64.8
G1 14.3
G2 M 20.8
S 64.1
G1 15.8
G2 M 20.1
S 64.7
G1 10.9
G2 M 24.3
S 52.2
G1 9.48
G2 M 38.2
S 63.0
G1 10.9
G2 M 26.1

5.2.1.2 The spike of H2A.X phosphorylation after treatment with G4 ligands

In the previous chapter, it became clear that the levels of DNA damage response proteins changed dramatically during differentiation. The ligands altered the proportion of cells in each phase of the cell cycle during differentiation, and treatment with PDS caused more cell death than in the undifferentiated state. Therefore, as shown on treatment with DNA damaging agents, it was likely that the G4 ligands would pose differing problems during differentiation compared to maintenance in the undifferentiated state.

To analyse whether the G4 ligands altered the level of γH2A.X at around 50 hours of differentiation during the epithelial mesenchymal transition, as discussed in the previous chapter, permeabilised flow cytometry was used. This was initially performed on cells treated with PDS throughout differentiation (Figure 71a). When cells were differentiated in the presence of PDS, the phosphorylation of H2A.X still occurred, suggesting that the cells were able to transition through the EMT. However, the level was in fact increased in the base population, indicative of DNA damage.

This upregulation of γH2A.X, believed to be associated with the EMT, also occurred in PhenDC3 treated cells but to a slightly lower extent (Figure 71b). There was no increase in the base level of H2A.X phosphorylation compared to mock-treated cells, suggesting that this ligand was acting in a different manner to PDS. However, PhenDC3 treated cells did reveal that the γH2A.X spike continued until 72 hours (Figure 71c), longer than the DMSO treated cells. However, since this was only performed once at these timepoints it is not possible to draw conclusions from this data.

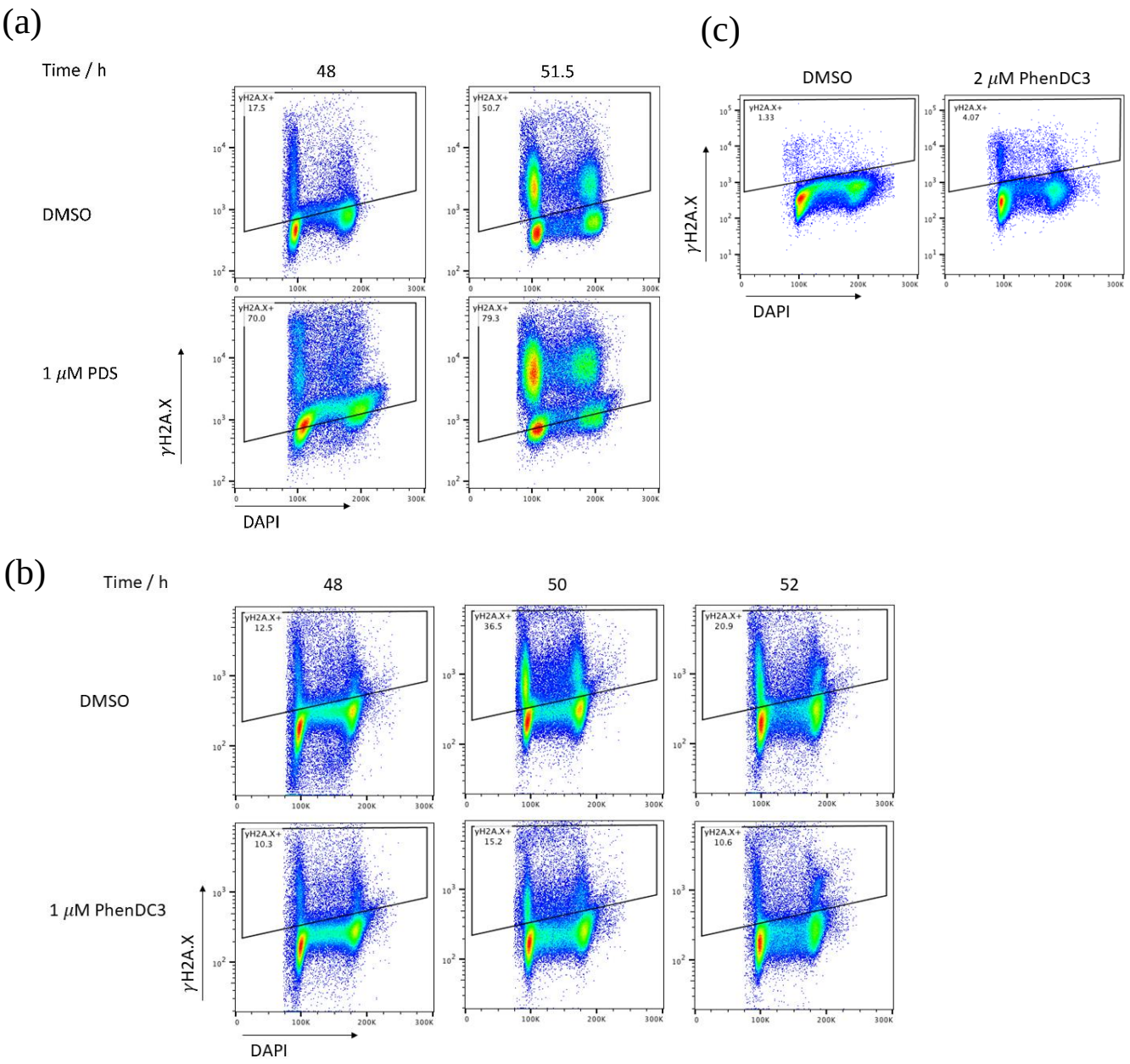


Figure 71. Analysis of the H2A.X spike of phosphorylation during differentiation in the presence of G4 ligands (a) in the presence of PDS (b) in the presence of PhenDC3 (c) PhenDC3 analysis at 72 hours. There is a baseline shift in the PDS treated cells, due to the damage induced non-specifically to the point of the EMT. Gating is relative to the DMSO treated spike of phosphorylation.

5.2.2 Gene expression changes during differentiation in the presence of G4 ligands

It was instantly clear when differentiating the cells in the presence of G4 ligands that not only were they perturbing the differentiation, but that PDS and PhenDC3 were doing so in distinct ways. PDS treated cells were larger by eye, especially at 24 hours, which was reflected in the 2D-cell cycle plots, there was more cell death and they looked similar to MMS treated cells. PhenDC3, however, caused the cells to metabolise more than untreated cells as seen by the

acidification of the medium and the cells remained as clumps rather than forming a mesenchymal sheet, this is discussed further in Section 5.2.2.2.

5.2.2.1 RNA changes during differentiation monitored by qPCR

RNA was collected from cells treated with PDS to understand whether the morphology and cell cycle differences described above were by changes to the gene expression profile. RNA was extracted from wildtype BOBSC cells grown with or without PDS in the undifferentiated state for three weeks and then differentiated for 72 hours and qPCR was performed (Figure 72a). This suggested that the cells treated with PDS were differentiating less efficiently than untreated cells, both in terms of the lower expression of the *SOX17* marker gene and in terms of maintaining higher expression of *T*. The higher level of *POU5F1* and *NANOG* suggested that these cells may be remaining in the undifferentiated state instead of committing to differentiation. The increased level of *T* mRNA is indicative of a more mesodermal fate, again suggesting that these cells are not fully committing to endodermal differentiation. This result implied that stabilising G4s in the genome could alter differentiation.

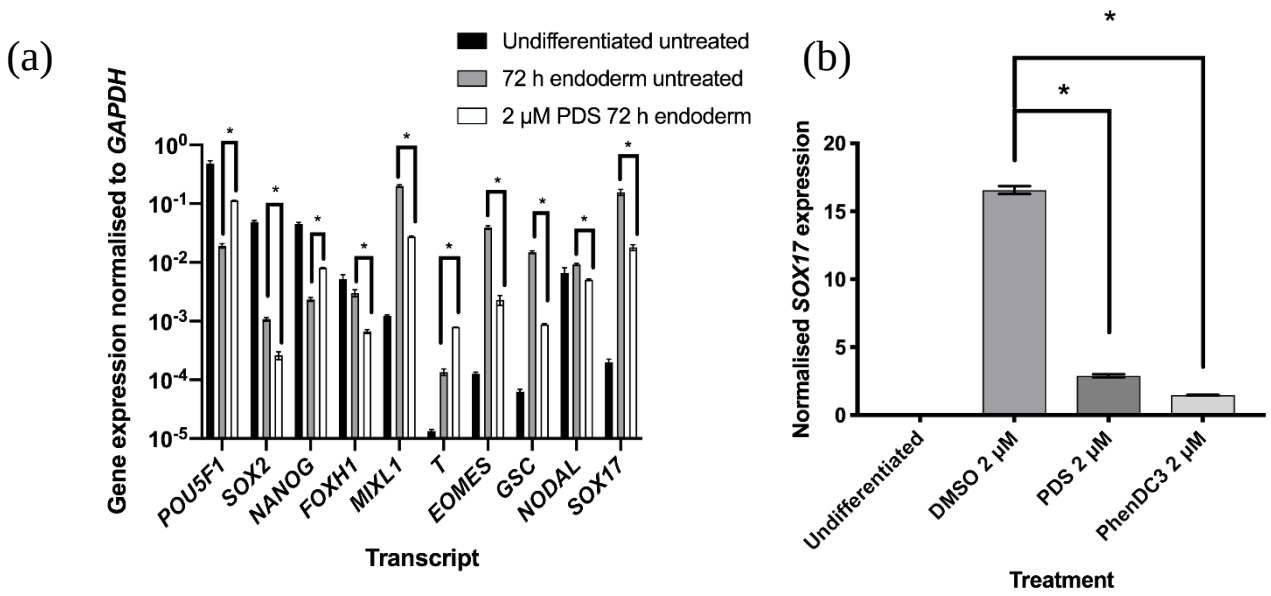


Figure 72. qPCR analysis of definitive endoderm differentiation in the presence of G4 ligands, n=3 technical replicates, mean±SD is plotted, multiple t tests were used to calculated significance (a) undifferentiated (black), differentiated (grey) and PDS treated (white), the first four transcripts should decrease during differentiation and the remaining increase, * denotes $p<0.05$ compared to the untreated differentiated samples (b) separate experiment to (a) the relative *SOX17* expression at 72 hours of differentiation in mock-treated samples compared to PDS and PhenDC3 treated samples normalised to the GEOMEAN. * denotes $p<0.05$, compared to differentiated untreated or DMSO treated cells.

This experiment was repeated using both PDS and PhenDC3: cells were differentiated in the presence of 2 µM PDS, PhenDC3 or DMSO and RNA was extracted for qPCR (Figure 72b). The level of *SOX17* decreased in the cells treated with ligand when normalised to the geometric mean of three control genes, *GAPDH*, *HPRT1* and *PBDG*. However, it was clear that PhenDC3 was altering the expression of *HPRT1*, which had been being used as a control gene. G4 ligands can affect the expression of control genes, and this was hard to predict. As this was hard to control for, further work was done using permeabilised flow cytometry and RNA sequencing in which this method of normalisation would not cause a problem. Since the experiment with PhenDC3 was only performed once, the significance cannot be tested but further experiments were used to verify these results.

5.2.2.2 Differentiation changes monitored using permeabilised flow cytometry

To get an initial picture of whether the ligands could perturb differentiation, the cells were grown in the undifferentiated state and differentiated in the presence of PDS (with reference to the schematic in Figure 11 (2)). This was so that they were grown in the presence of ligands for the longest period of time and an effect would have been likely to be seen. Cells were maintained in 2 µM PDS for over two weeks and then differentiated in the presence of PDS. They were collected for permeabilised flow cytometry every 24 hours during the differentiation process and stained with αSOX17 and αEOMES antibodies (Figure 73). Cells differentiated in the presence of PDS exhibited more cell death than untreated cells. The cell cycle of the treated cells was perturbed such that at 24 hours there was a large G2/M block, as seen previously, and by 72 hours the cells were not cycling in a normal manner. The expression of the mesendoderm marker EOMES was upregulated at 24 and 48 hours in both the ligand and untreated conditions but was downregulated to a greater extent at 72 hours when treated with PDS, possibly due to the sickness of the cells at this point. While the differentiation in this experiment had not been particularly efficient, it was clear that SOX17 was not being expressed at all in PDS treated cells at 48 hours and the level had not increased much by 72 hours. The G2/M phase cell cycle block early in differentiation was reminiscent of that seen with MMS treatment discussed in Results II. This increase in proportion of G2/M cells was clearly visible from the increase in size seen down a microscope at 24 hours. While this was at a high concentration of PDS, the same trend was seen at all concentrations and was likely to be impacting differentiation severely.

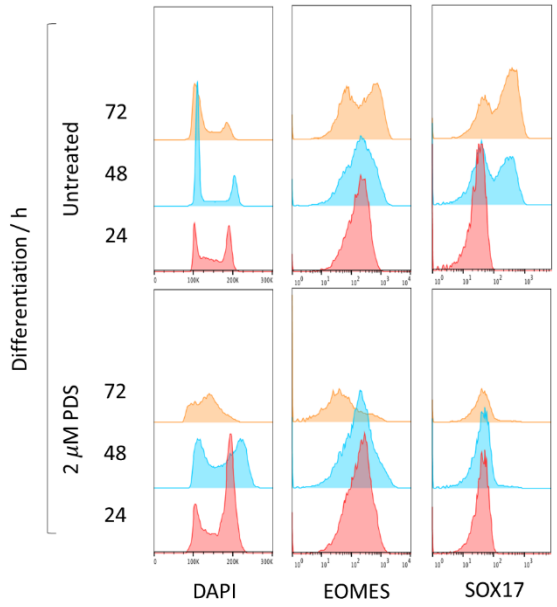


Figure 73. Analysis of the differentiation of BOBSC cells cultured in the undifferentiated state with PDS before being differentiated in the presence of this ligand. Cells were cultured in the undifferentiated state with 2 μM PDS for over two weeks before being differentiated in the presence of the ligand and differentiation was analysed every 24 hours for DNA content, EOMES and SOX17 expression.

Once it was clear that culturing cells in PDS in both the undifferentiated state and during differentiation prevented efficient SOX17 upregulation I wanted to ask the question, does PDS perturb the ground state and does it have a permanent and heritable impact on ground state genes expression and the ability to differentiate? This had been addressed previously in the mouse cells (with reference to the schematic in Figure 11 (1) with continued differentiation). Cells grown on the drug for 2.5 weeks were cultured without the drug for 2.5 weeks and then differentiated without PDS. The proportion of SOX17 expressing cells was quantified and compared to the untreated cells (Figure 74). This experiment clearly demonstrated that if PDS was not present at the time of differentiation then there was very little or no effect on the differentiation. There were also no changes to the morphology, or any increased cell death by observation. However, the cells maintained in PDS differentiated only half as efficiently.

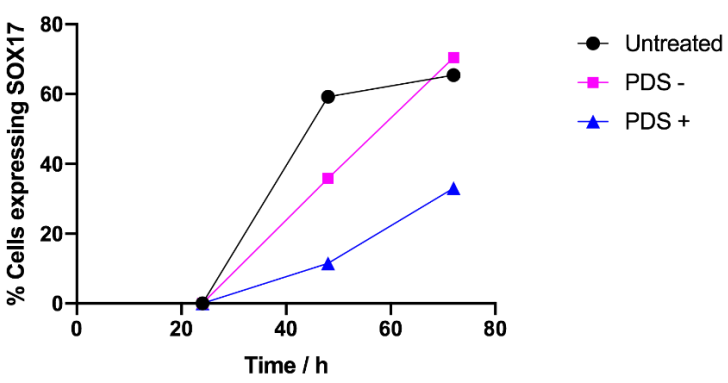


Figure 74. Analysis of whether G4 ligands in the undifferentiated state could alter differentiation without the ligands being present during this process. Cells were cultured with 2 µM PDS for 2.5 weeks and then either washed out of drug and cultured for a further 2.5 weeks without the drug before being differentiated without the drug (PDS -). Cells kept in culture with PDS throughout this time were used as a control (PDS +) and without any drug (untreated).

The next question to address was whether the presence of G4 ligands during differentiation alone could alter the outcome. The differentiation was repeated in the presence of 2 µM PDS using cells that had not been grown on drugs prior to differentiation and cells were collected at 72 hours for permeabilised flow cytometry (Figure 75). This showed a similar picture to the cells grown in 2 µM PDS prior to differentiation and differentiated in the ligand: SOX17 was not upregulated to the same extent as in the untreated cells (Figure 75c).

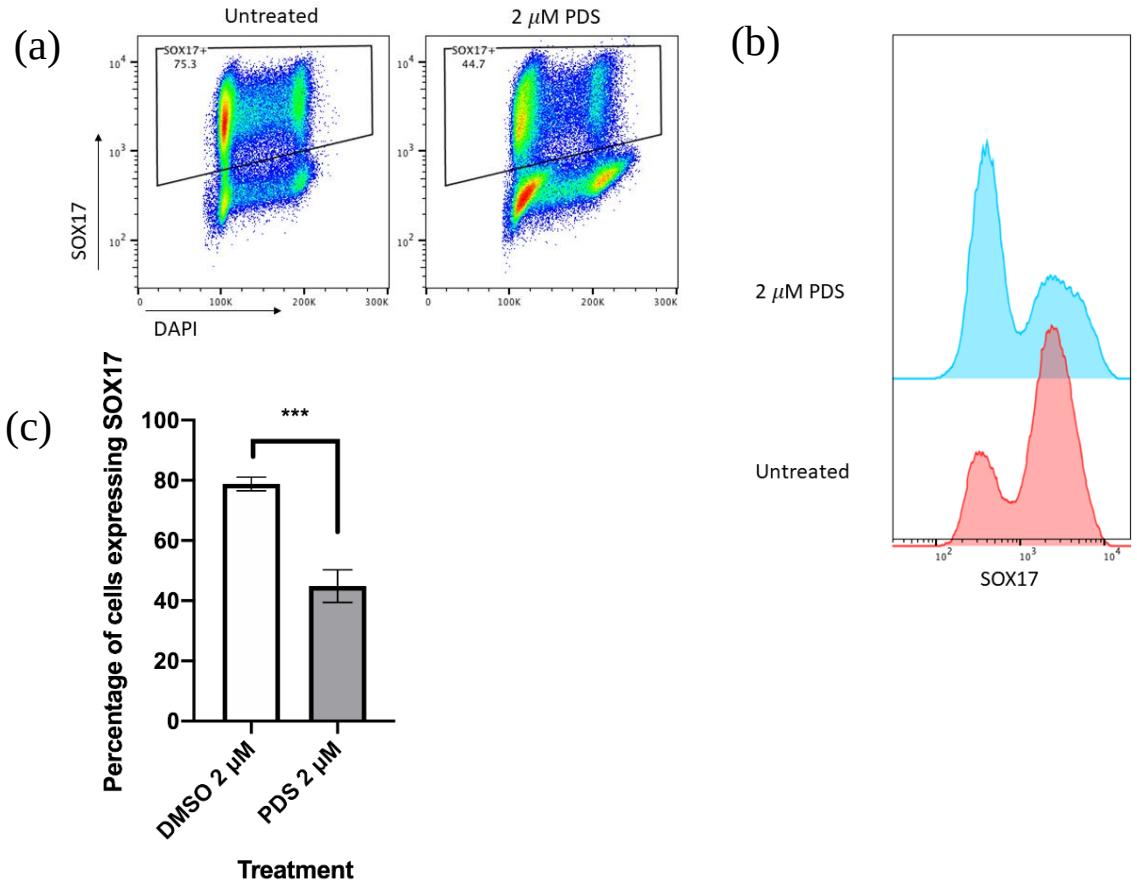


Figure 75. Flow cytometry analysis of cells cultured with 2 μM PDS from the onset of differentiation, throughout differentiation. Cells were collected at 72 hours to analyse the proportion of SOX17 positive cells with or without treatment (a) a plot showing SOX17 positive cells in relation to the DNA content (b) histograms showing the expression of SOX17 (c) replicate experiments showing the proportion of cells expressing SOX17 at 72 hours n=7, paired t test p=0.0001, mean±SEM are plotted.

From these preliminary experiments it was evident that if PDS was not present during endoderm differentiation then there was no effect on the SOX17 positive population of cells. All further experiments were performed without growing the cells in ligand prior to differentiation but adding the G4 ligand at the initiation of differentiation, with the ligand being replaced every 24 hours when the differentiation media was changed. This also made the experiments easier to interpret with respect to the pathways of perturbed gene expression changes.

PDS exposure during differentiation decreased the efficiency of differentiation but it was also causing some damage and death to the cell as well as altering the cell cycle, comparable to the effect experienced in MMS treated cells in Results II. In order to understand whether PDS was causing a G4-specific effect that was independent of inducing damage, the G4 ligand PhenDC3 was used. This ligand has been shown to stabilise G4s in DT40 at the concentrations used in this thesis (Guilbaud et al., 2017). Cells were differentiated in the presence of 2 μM PhenDC3 and collected every 24 hours for permeabilised flow cytometry (Figure 76a and b). EOMES was upregulated at 24 hours to a similar extent in both the untreated and PhenDC3 treated cells

but SOX17 was upregulated to a lesser extent at both 48 and 72 hours in the PhenDC3 treated cells (Figure 76c). As is expected during endoderm differentiation, the proportion of cells in G1, as seen with DAPI staining, increased during differentiation which was in contrast to the PDS treated cells. Another interesting point to note was that cells differentiated in the presence of PDS exhibited more cell death than the untreated cells whereas PhenDC3 treated cells metabolised to a greater extent than untreated cells, shown with the greater acidification of the medium. Looking down the microscope, there were also some small phenotypic differences between the conditions including PhenDC3 treated cells remaining as individual colonies rather than forming a monolayer as in the unperturbed differentiation.

Higher concentrations of PhenDC3 did not cause increased cell death over 72 hours compared to the PDS treatment and therefore it was possible to analyse the dose response of the efficiency of differentiation to this G4 ligand over a range of concentrations (Figure 77). There was a clear decrease in the percentage of SOX17 expressing cells as the PhenDC3 concentration increased, showing a dose response to this ligand.

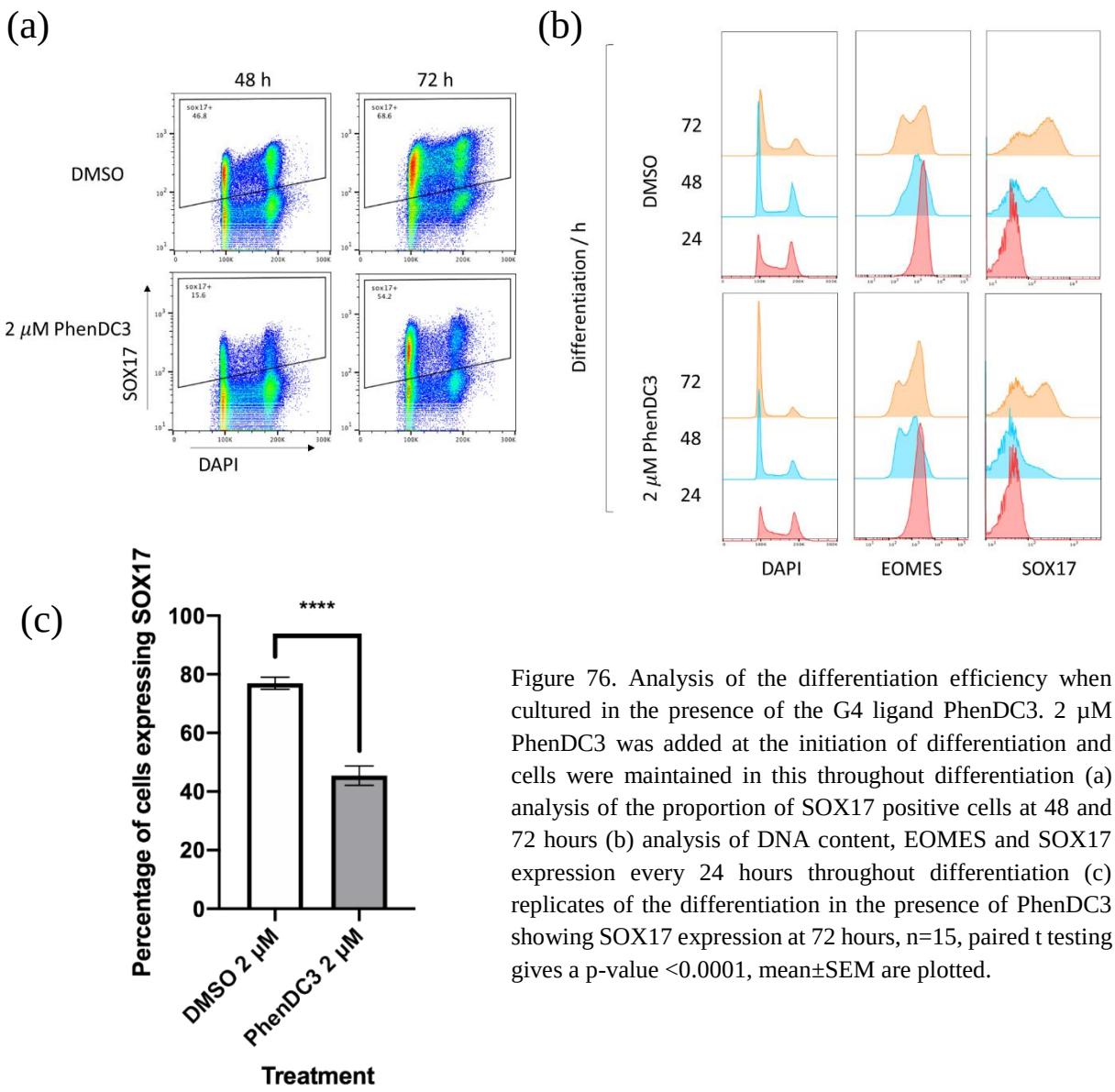


Figure 76. Analysis of the differentiation efficiency when cultured in the presence of the G4 ligand PhenDC3. 2 µM PhenDC3 was added at the initiation of differentiation and cells were maintained in this throughout differentiation (a) analysis of the proportion of SOX17 positive cells at 48 and 72 hours (b) analysis of DNA content, EOMES and SOX17 expression every 24 hours throughout differentiation (c) replicates of the differentiation in the presence of PhenDC3 showing SOX17 expression at 72 hours, n=15, paired t testing gives a p-value <0.0001, mean±SEM are plotted.

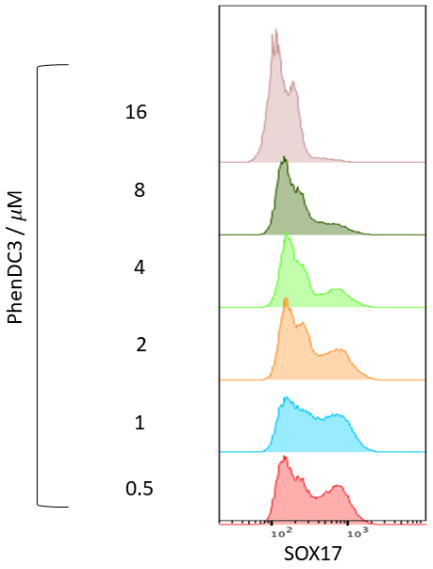


Figure 77. The proportion of SOX17 expressing cells at 72 hours of definitive endoderm differentiation with cells cultured in increasing concentrations of PhenDC3.

5.2.2.3 Differences between the G4 ligands

These experiments showed that there were many differences between the cells treated with these two G4-binding ligands. While this made the observations hard to interpret, it was interesting that they were exhibiting different effects. As the PDS caused many changes that were comparable to the DNA damaging agents, these experiments were repeated in the *TP53*$^{-/-}$ cells described in Results II.

5.2.3 Comparison of G4 ligand effects with DNA damaging agents

5.2.3.1 The role of p53 in differentiation in the presence of G4 ligands

In Results II it was shown that *TP53*$^{-/-}$ cells did not inhibit endoderm differentiation in response to DNA damage as wildtype BOBSC cells. In order to understand whether this was also the case with the ligands, these experiments were repeated using PDS and PhenDC3. The G4 ligands were added to *TP53*$^{-/-}$ cells throughout differentiation and compared to wildtype treated cells (Figure 78a and b). PDS treated *TP53*$^{-/-}$ cells differentiated with a much greater efficiency compared to the wildtype treated cells: the knockout of *TP53*$^{-/-}$ at least partially compensated for the addition of the drug. PDS appears to act, at least in part, by inducing DNA damage. However, the PhenDC3 treated cells showed a different picture compared to the MMS, UV and PDS treated *TP53*$^{-/-}$ cells: the PhenDC3 treated *TP53*$^{-/-}$ cells showed the same perturbation of differentiation as wildtype cells. This provided further evidence that PhenDC3 specifically was not simply preventing differentiation by inducing DNA damage.

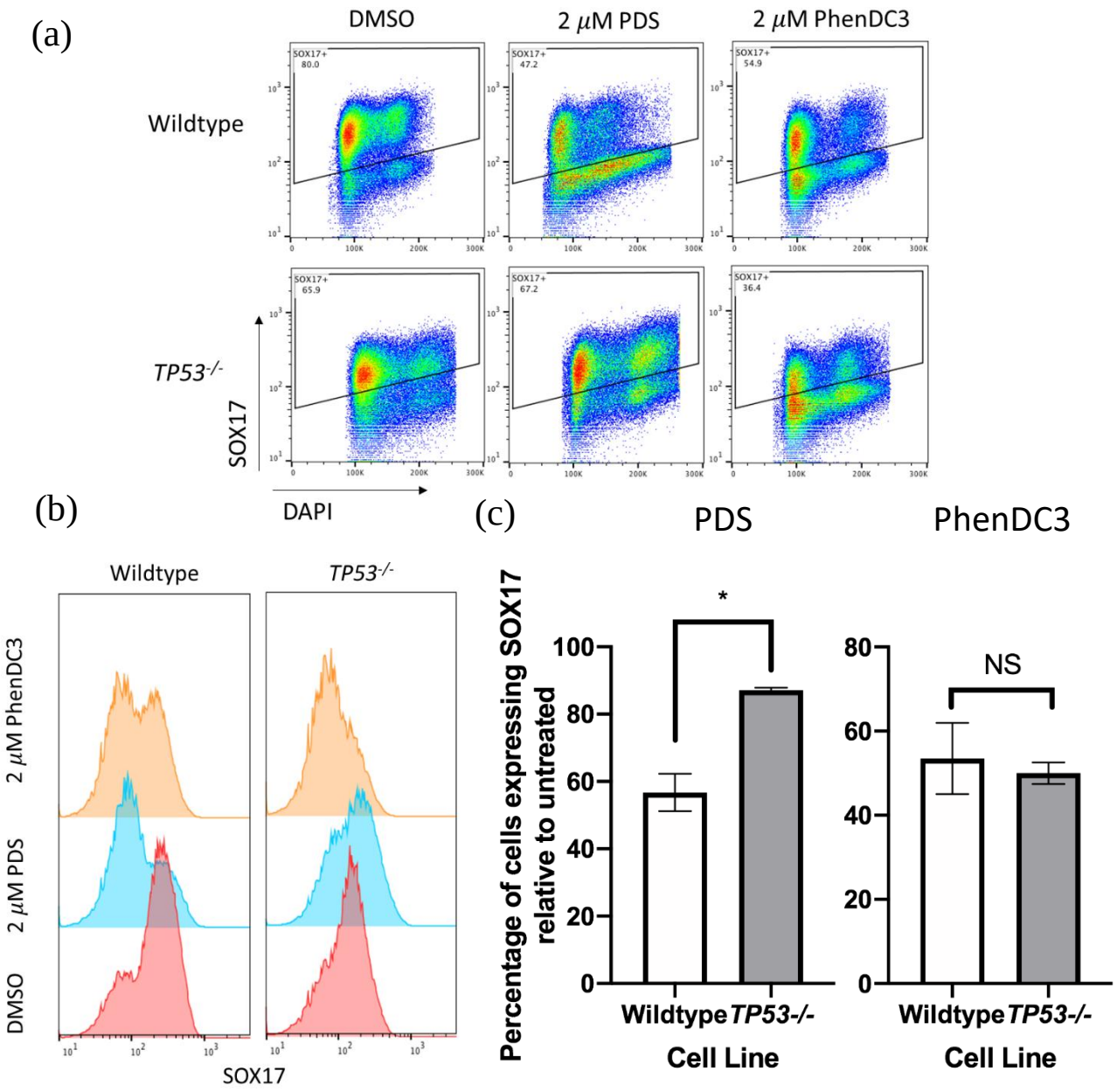


Figure 78. Analysis of the role of p53 in differentiation in the presence of the G4 ligands by using the *TP53*$^{-/-}$ cell line as in Results II (a) plots to show the percentage of SOX17 positive cells at 72 hours in each condition (b) histograms showing the expression of SOX17 at 72 hours (c) replicate experiments showing the expression of SOX17 at 72 hours in the wildtype vs *TP53*$^{-/-}$ cell lines with PDS (left) or PhenDC3 (right) compared to untreated. n=2 for PDS, unpaired t test p=0.0326, n=5 for PhenDC3 with no significance found, mean±SEM are plotted.

5.2.3.2 Monitoring a delay in differentiation with G4 ligands

In order to see whether PDS was simply slowing differentiation down and the cells would require longer to catch up with SOX17 expression, the differentiation was continued for a further 24 hours, until 96 hours. However, this experiment caused so much cell death that there were not enough cells to analyse for permeabilised flow cytometry. This suggested that the cells would not differentiate in the presence of PDS, regardless of the length of time. To see whether removing the drug could reverse the effect on differentiation, cells were differentiated for 72

hours in the presence of PDS and then the PDS was removed for 24 hours and the cells were collected for flow cytometry at 96 hours (Figure 79a). Although there were not many cells left to analyse at this timepoint, it seemed that removing the drug for the final 24 hours did slightly increase the number of cells expressing SOX17 but since there were not many cells left alive it was not clear.

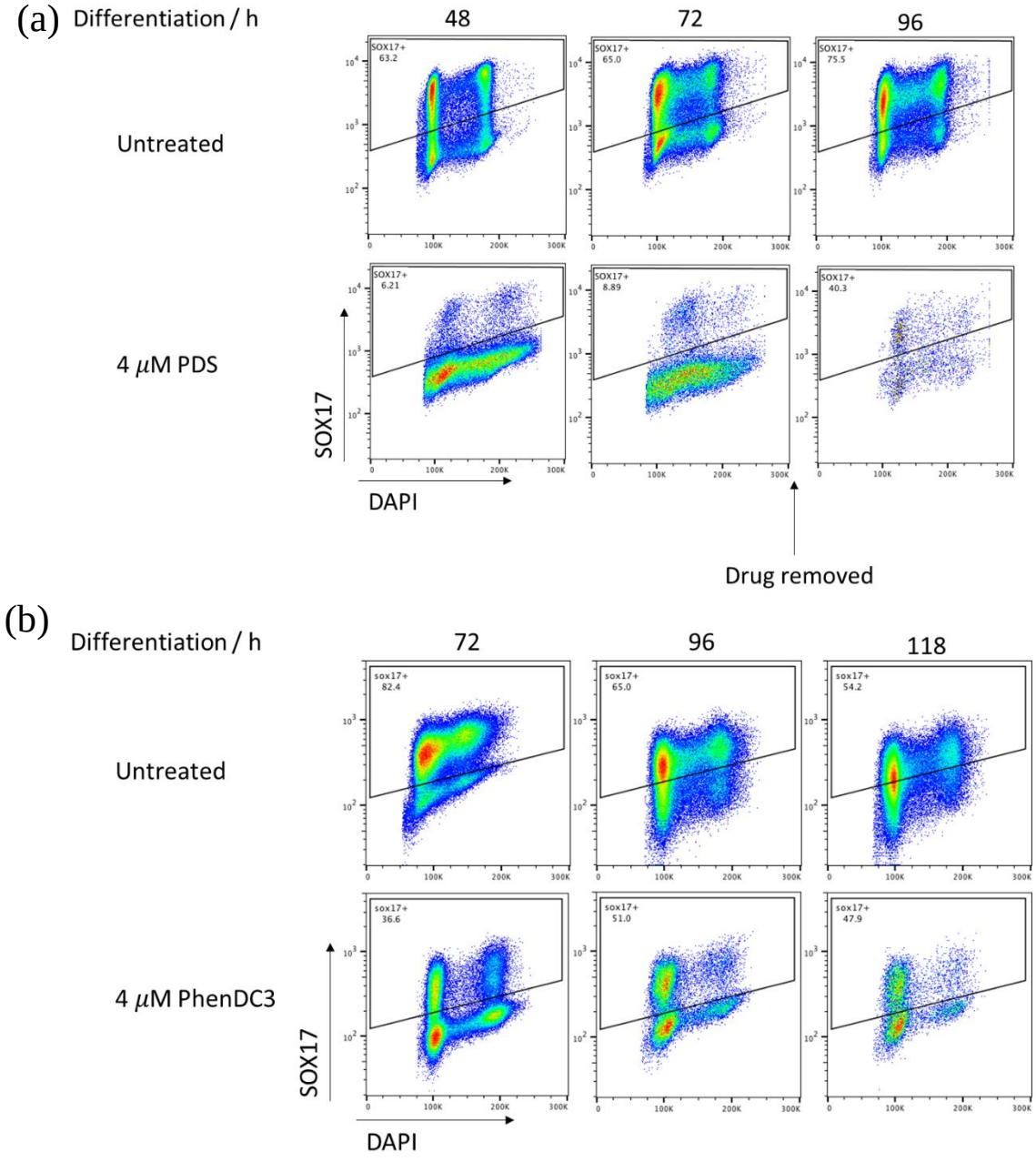


Figure 79. Differentiating the cells for over 72 hours (a) PDS was added for the first 72 hours and then removed for the final 24 hours before the cells were collected at 96 hours to analyse the proportion of SOX17 positive cells (b) cells were cultured in the presence of PhenDC3 for an extra 24 or 46 hours and analysed for the proportion of SOX17 positive cells.

PDS treatment caused a high level of cell death and therefore it is hard to maintain the differentiating cells for over 72 hours, but this was not the case for PhenDC3. PhenDC3 treated cells were differentiated for longer than 72 hours to monitor whether the differentiation could catch up (Figure 79b). This showed that an extra 46 hours of differentiation did not allow the PhenDC3 treated cells to express SOX17 to as high a level to the untreated, there was still a distinct population not expressing SOX17. This suggested that the cells would not differentiate even given the extra time. Again, suggesting that PhenDC3 perturbs a pathway early on in the differentiation process. However, the untreated cells got too confluent over this time so the proportion of cells expressing SOX17 decreased.

(a)

	Median number of cell cycles	
2 μM Treatment	**Wildtype**	***TP53*$^{-/-}$**
DMSO	5.1	4.5
PDS	4.3	4.1
PhenDC3	4.6	3.9

(b)

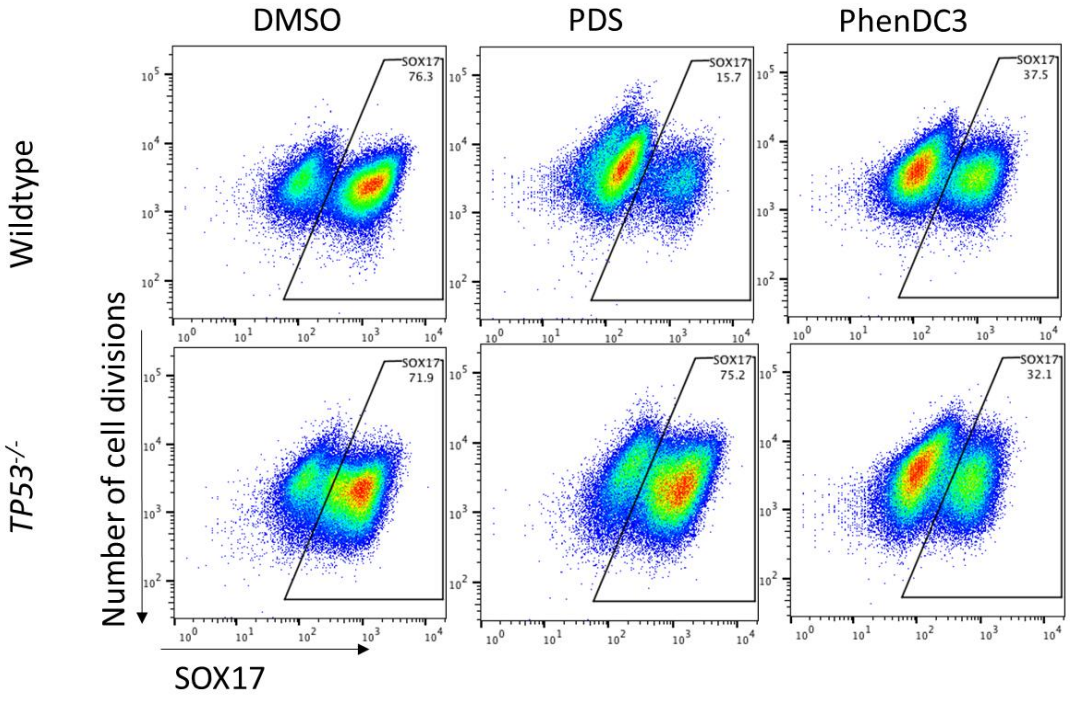


Figure 80. Analysis of the number of cell divisions gone through during differentiation in the wildtype and *TP53*$^{-/-}$ cell lines with each treatment at 72 hours (a) a table to show the median number of divisions (b) the proportion of SOX17 expressing cells in each condition in relation to the number of cell divisions. Fluorescence decreases by half on each cell division. One experiment shown.

The CellTrace™ approach, as used in Section 4.2.6.2, was also used to monitor the number of cell divisions in cells treated with 2 µM PDS or PhenDC3 in wildtype and *TP53*$^{-/-}$ cells (Figure 80). As in the MMS treated cells there was very little difference between the DMSO and PhenDC3 or PDS treated cells in terms of the number of cell divisions, suggesting it was not a lack of cell division that prevented differentiation in this case. In all cases it was clear that the cells had gone through enough rounds of division for SOX17 to be expressed. However, this experiment should be repeated to assess whether these changes are significant.

5.2.3.3 Activity of PhenDC3 on the cells throughout differentiation

In the case of MMS addition, the results showed that there was a key window in which this had the greatest effect on altering the differentiation: this may be a DNA damage specific response. To ask whether the PhenDC3, which was not thought to be inducing DNA damage, was acting in a similar manner, the experiment was repeated by differentiating the cells in PhenDC3 for either 0-24, 24-48 or 48-72 hours during differentiation and analysing the differentiation at 72 hours (Figure 81). Despite PhenDC3 decreasing the proportion of cells committing to definitive endoderm differentiation at all times, the biggest decrease was seen if cells were treated for the first 24 hours of differentiation. This suggested that the most crucial time for having PhenDC3 present was in the first 24 hours of the differentiation, consistent with it not causing general DNA damage.

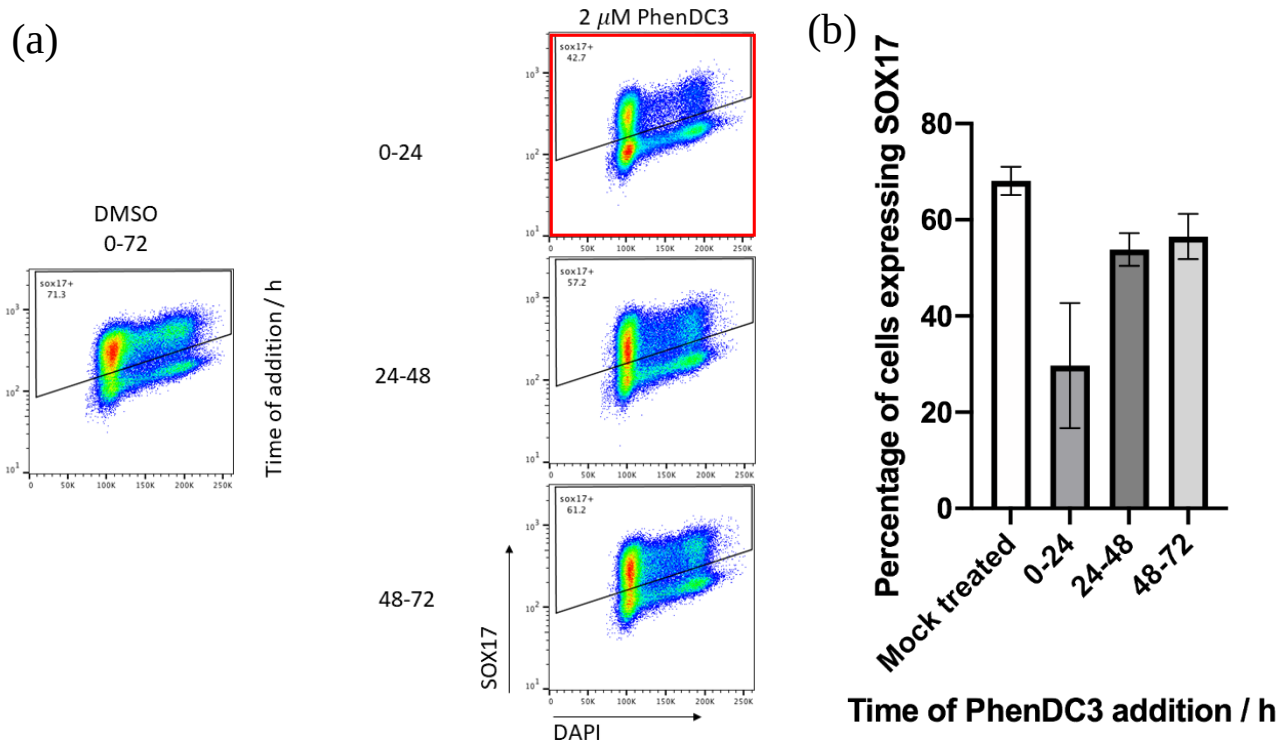


Figure 81. PhenDC3 was added for 24-hour windows during differentiation to understand whether there was a critical point for the addition on altering the course of differentiation. The percentage of SOX17 expressing cells was monitored at 72 hours and compared to the DMSO treated cells (a) one experiment (b) n=2, no significance using paired t test, mean±SEM are plotted.

In a similar manner to the DNA damaging agents experiment in which the exposure was made after 72 hours of differentiation and the SOX17 positive population was monitored, Section 4.2.5, this was performed using the G4 ligands. This experiment allowed an understanding of whether the expression of SOX17 could only be altered when the cells were differentiating, as with the DNA damaging agents, or if they can be altered once the cells had differentiated (Figure 82). As seen for the DNA damage, these results show that PhenDC3 and PDS only affected the differentiation if they were added during the differentiation process, and when SOX17 was stably expressed, the cells did not stop expressing it regardless of the treatment. This suggested that PhenDC3 and possibly PDS were altering a pathway early in differentiation.

To understand the gene expression changes caused by these ligands, RNA sequencing was performed throughout differentiation.

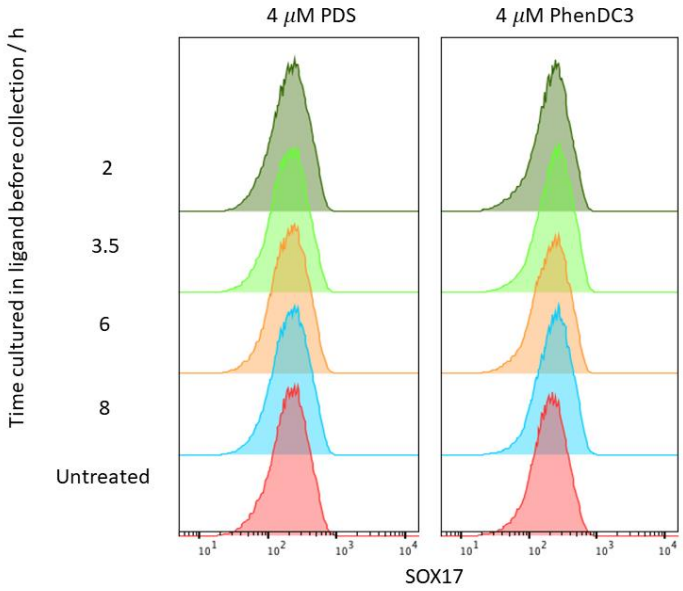


Figure 82. Analysis of whether the addition of G4 ligands after 72 hours alters differentiation. Cells were cultured in 4 μM ligand for between two and eight hours before collection and analysis of the SOX17 expressing cells at 80 hours of differentiation. PDS (left panel) and PhenDC3 (right panel).

5.3 RNA sequencing analysis of G4 ligand treated cells in the undifferentiated state and during differentiation

In order to get a detailed analysis of the way in which the G4 ligands were perturbing differentiation, RNA sequencing was performed. Cells were differentiated in triplicate in the presence of 2 μM PDS, 2 μM PhenDC3 or the equivalent volume of DMSO and RNA was collected every 24 hours. This was performed in parallel with maintaining the cells in the undifferentiated state with the drugs to understand which effects were specific to differentiation, as discussed in Section 5.1.4. Some compromises were made with respect to the density and time of collection as there was no perfect solution for collecting RNA for sequencing; although initial data using the BOBSC knockout cells showed that there was a large 'batch effect' on the transcriptome likely to be due to the confluence of the cells, Section 6.3.1, it is also known that the efficiency of the mesendoderm pathway is affected by the confluence of the cells during the first 24 hours and this had been noted in my experiments (Kempf et al., 2016). Therefore, cells were plates at the same density and colony size and each timepoint was collected 24 hours after the previous timepoint, therefore not all at the same time. After the RNA was extracted and the quality was checked, RNA libraries were created.

The principle component analysis (PCA) plot was generated by Alastair Crisp (Figure 83a and b). This shows a good clustering of the undifferentiated cells at each timepoint with each of the drugs on the left side of the plot. However, as discussed in Section 6.3.1, it was clear that the undifferentiated cells were still experiencing a 'batch effect' likely to be due to the confluency of the cells in the well. These undifferentiated, ligand-treated cells did not show any major

differences compared to the DMSO treated cells in their expression profiles as seen in the PCA. What was very clear was that cells differentiating to definitive endoderm showed very large gene expression changes during differentiation, compared to the undifferentiated cells, as expected. The number of gene expression changes during DMSO differentiation was compared to the undifferentiated 24-hour DMSO treated cells and the Venn diagram is shown in Figure 84.

However, the PhenDC3 treated cells had a very different gene expression program compared with the vehicle-treated cells and this was very clear from 48 hours into the differentiation. This was less dramatic in the PDS treated cells which showed only 376 genes significantly altered between its expression and the mock-treated cells at 72 hours compared to 2576 in PhenDC3 treated cells and over half of the PDS differentially expressed genes were shared with PhenDC3 (Figure 85), genes considered to be differentially expressed are discussed in the materials and methods section. The PCA was very different between the two drugs supporting the conclusion that the drugs were acting through different mechanisms. This gave further support to the idea that PhenDC3 did not cause DNA damage and upregulate the DDR. Deregulated transcripts were compared to those deregulated at 72 hours in the MMS treated cells, normalised to untreated and *TP53*$^{-/-}$ MMS treated cells (Figure 86a and b). This clearly shows that the majority of transcripts associated with PDS treatment, were also altered on the addition of MMS, whereas the majority of those deregulated on treatment with PhenDC3 were not common to MMS treatment. I hypothesise that the changes not associated with MMS are likely to be specific to G4 formation.

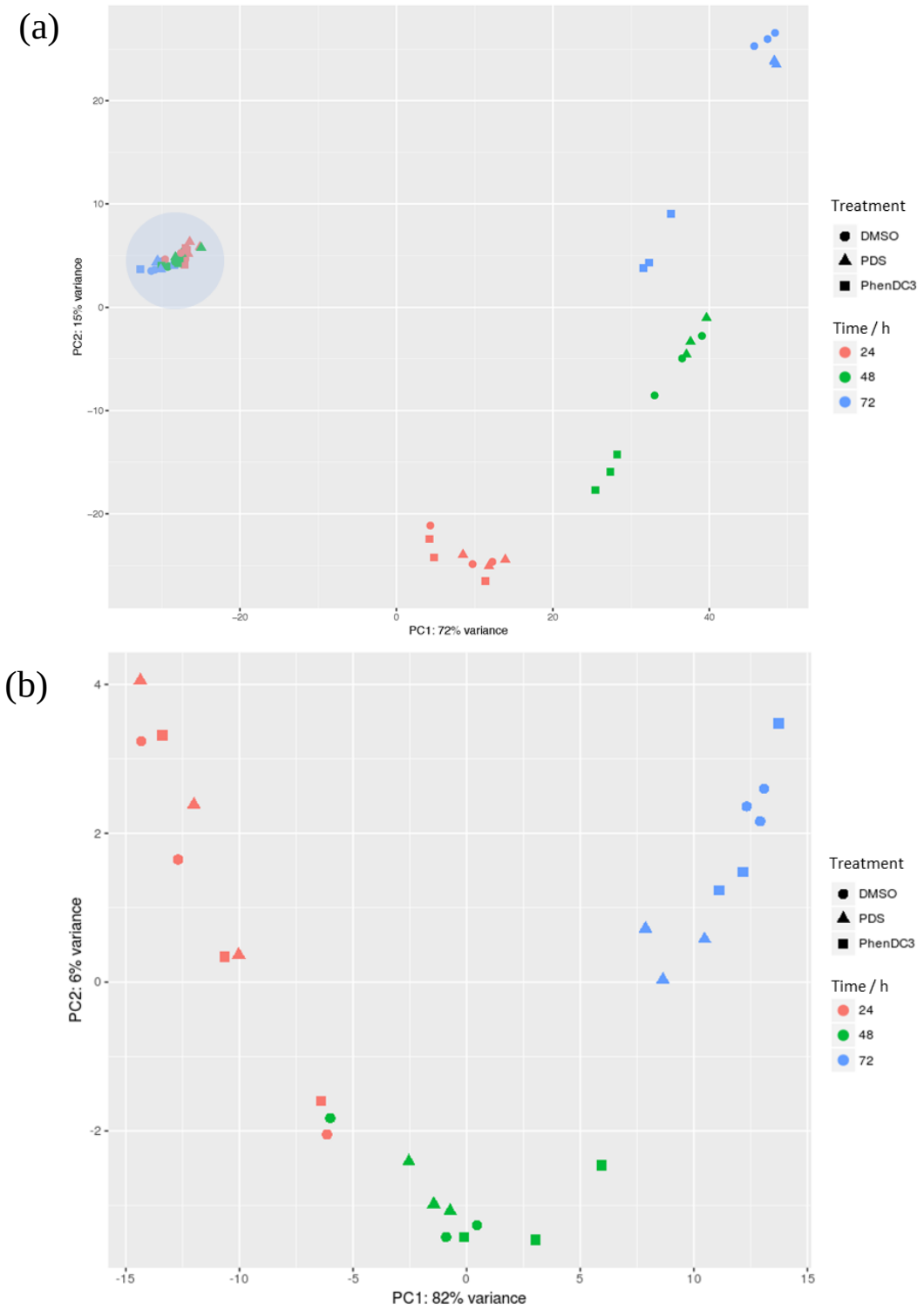


Figure 83. The PCA generated to show the maximum variance between all of the samples in the G4 ligand differentiation RNA sequencing experiment. (a) All samples: undifferentiated cells are clustered together on the left-hand side of the plot (shown in the blue circle), whereas differentiating samples fall to the right-hand side. DMSO mock-treated cells are shown as circles, PDS as triangles and PhenDC3 as squares. Samples collected at 24 hours are shown in red, 48 hours in green and 72 hours in blue. (b) Only undifferentiated samples, labelled as (a).

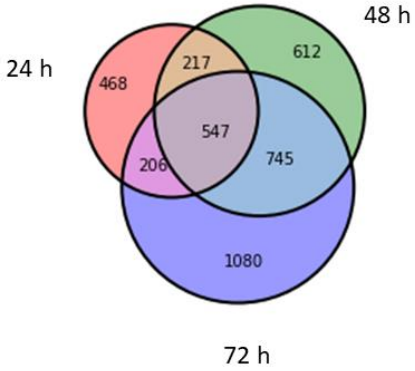


Figure 84. A Venn diagram to show the number of genes significantly changing expression during unperturbed differentiation compared to the undifferentiated 24-hour mock-treated cells. The number of genes which change expression increase every 24 hours.

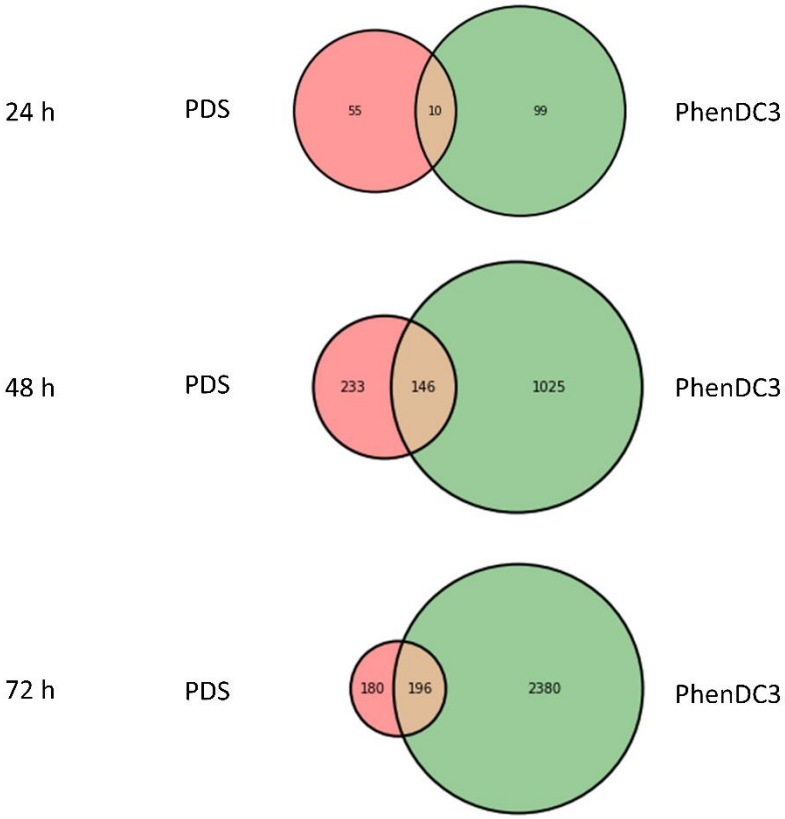


Figure 85. Venn diagrams to show the comparison of genes deregulated in PDS and PhenDC3 treated cells during differentiation. Differentiation was compared to the DMSO treatment at each time point. Venn diagrams are not to scale between conditions.

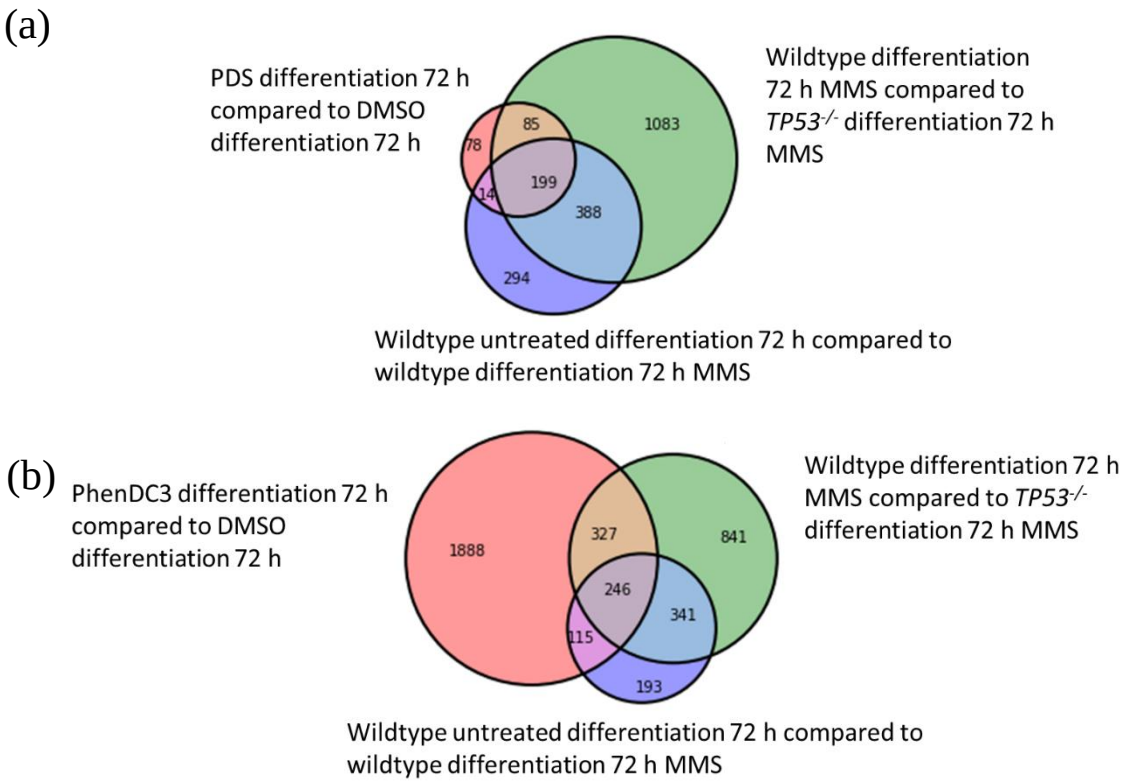


Figure 86. A comparison of ligand treatment with MMS treatment in Results II. The genes with altered expression at 72 hours of differentiation when treated with PDS (a) or PhenDC3 (b) were compared to DMSO treatment. These genes were then compared to genes with deregulated expression in wildtype MMS treated cells compared to wildtype untreated cells, as well as genes with altered expression when comparing the wildtype cell line treated with MMS and the *TP53*$^{-/-}$ cell line treated with MMS.

5.3.1 Gene expression changes in the undifferentiated state with PDS and PhenDC3

In order to understand whether the G4 ligands were perturbing the undifferentiated state in the BOBSC cells, the deregulated genes were compared with both ligands. The undifferentiated PDS and PhenDC3 treated cells grown with the drug for 72 hours were compared with undifferentiated cells grown in DMSO. This was the most direct comparison to minimise and effects of confluence on separate days. The deregulated genes in both conditions were compared between the drugs and this is shown in the Venn diagram in Figure 87. This clearly shows that only 49 genes were deregulated in the PDS treated cells, and 16 in the PhenDC3 treated cells. There were 150-fold fewer deregulated genes in the PhenDC3 treated undifferentiated cells compared to the differentiated cells, highlighting the difference in response to this ligand between the two cell states. In the PDS treatment this difference was less pronounced, the difference further agreed with the findings that these ligands were not perturbing the gene expression in the undifferentiated state. The overlap between the drug treated cells was only six genes: *RNF170*, *AMH*, *NRBP2*, *MEF2C*, *GAN* and *ANXA1*.

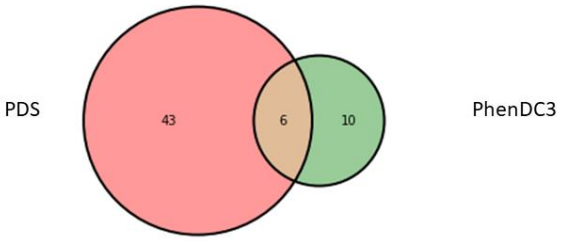


Figure 87. A Venn diagram to show the number and overlap of genes with deregulated expression in the undifferentiated state after 72 hours of treatment with either PDS (left) or PhenDC3 (right).

The ligands were causing much greater expression changes during differentiation. This may be due to the fact that there are such vast changes during differentiation and therefore more pathways to perturb. It may also be that when cultured in the pluripotent state, cells often lose adherence when they lose pluripotency, therefore cells that have lost expression of key genes and floated off the dish could not be counted. The deregulated genes are compared to the knockout cell lines in Results IV. The GO term analysis for the 72-hour undifferentiated cells treated with G4 ligands is shown in Table 20.

In the PDS treated cells there were a number of highly significant gene ontology functions associated with this gene list. These were mainly related to development, suggesting that PDS may be causing the cells to differentiate or lose pluripotency in the undifferentiated state. However, with PhenDC3 treatment, there were not many significant pathways and almost all of them were only significant due to the deregulated expression of three genes: *ANXA1*, *IDO1* and *MEF2C*. *ANXA1* encodes Annexin A1 which is a membrane-localised protein which binds phospholipids. This protein has anti-inflammatory activity and loss of function is detected in many tumours. *IDO1* encodes indoleamine 2,3-dioxygenase which is a heme enzyme catalysing tryptophan catabolism. *MEF2C* encodes myocyte enhancer factor 2C which is a MADS box transcription enhancer factor 2. It is involved in cardiac morphogenesis and myogenesis. Analysis of the overlap of PDS and PhenDC3 was skewed due to lack of genes but consisted of *RNF170, AMH, NRBP2, MEF2C, GAN* and *ANXA1*.

GO term	Description	Count	Adjusted p-value
PDS			
GO:0032835	Glomerulus development	5/61	0.00024
GO:0048869	Cellular developmental process	25/4270	0.00024
GO:0030154	Cell differentiation	25/4088	0.00024
GO:0035850	Epithelial cell differentiation involved in kidney development	4/45	0.00107
GO:0070848	Response to growth factor	10/713	0.00107
GO:0071363	Cellular response to growth factor stimulus	10/685	0.00107
GO:0070431	Nucleotide-binding oligomerisation domain containing 2 signaling pathway	3/13	0.00107
GO:0001655	Urogenital system development	7/317	0.0018
GO:0061005	Cell differentiation involved in kidney development	4/55	0.00188
GO:0033209	Tumor necrosis factor- mediated signaling pathway	5/127	0.00224
PhenDC3			
GO:2000108	Positive regulation of leukocyte apoptotic process	3/29	0.00167
GO:0033034	Positive regulation of myeloid cell apoptotic process	2/8	0.0105
GO:2000106	Regulation of leukocyte apoptotic process	3/83	0.0137
GO:0071887	Leukocyte apoptotic process	3/100	0.0179
GO:0006925	Inflammatory cell apoptotic process	2/20	0.046
GO:0033028	Myeloid cell apoptotic process	2/32	0.046
GO:0033031	Regulation of myeloid cell apoptotic process	2/28	0.046
GO:0002828	Regulation of type 2 immune response	2/30	0.046
GO:0036211	Protein modification process	9/4097	0.047
GO:0006464	Cellular protein modification process	9/4097	0.047

Table 20. The ten most significantly enriched gene ontology (GO) terms associated with PDS compared to DMSO and PhenDC3 compared to DMSO after 72 hours of culture with the ligands in the undifferentiated state.

5.3.2 Gene expression changes during differentiation with G4 ligands

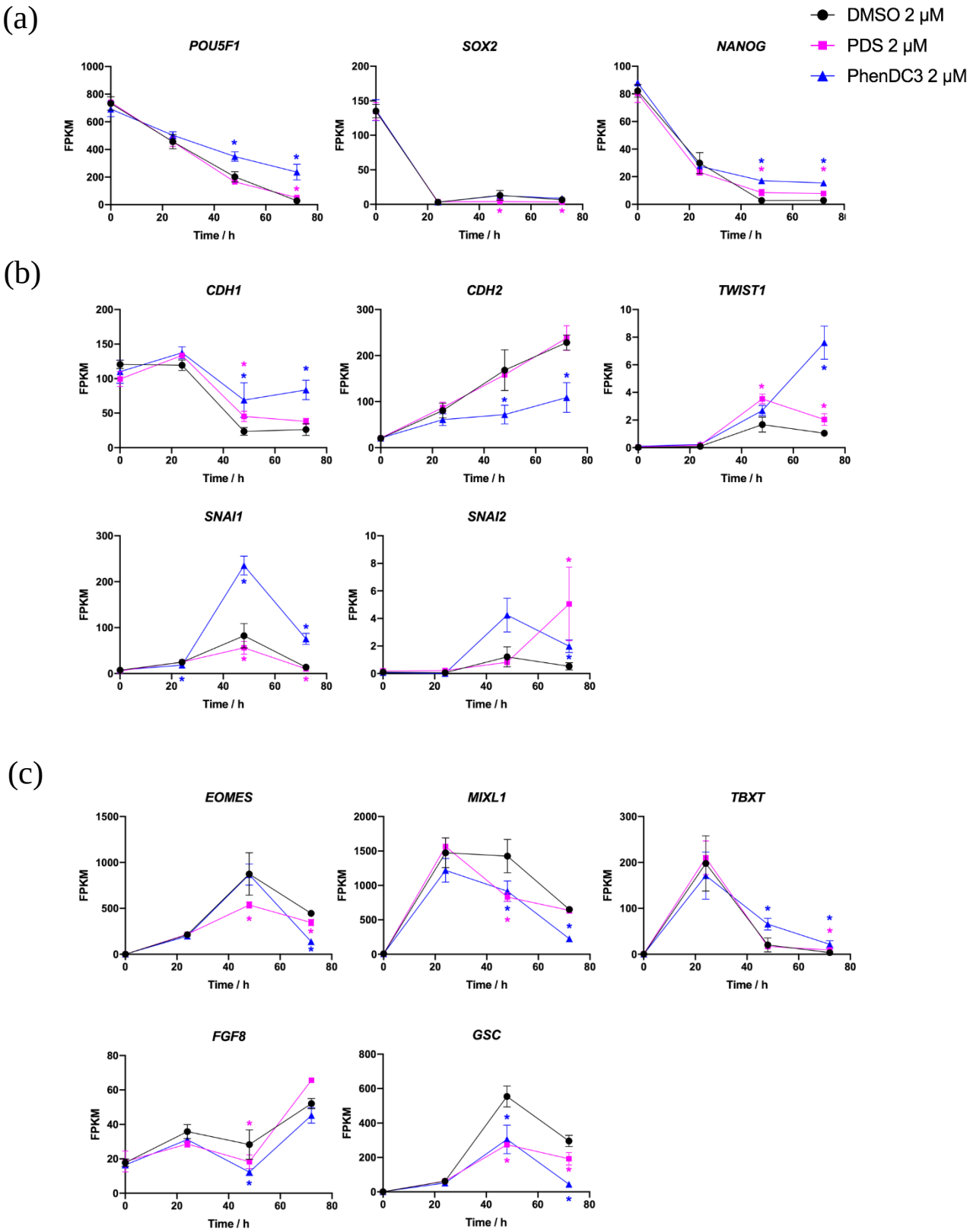
Firstly, the main genes that are known to change expression during the differentiation process were analysed from the sequencing data (Teo et al., 2011) (Figure 88). The pluripotency genes, *NANOG*, *POU5F1* and *SOX2*, decreased expression throughout the differentiation process, as expected (Figure 88a). *SOX2* was switched off in the first 24 hours and this was the case with both ligands as well as the mock treatment. *NANOG* and expression was switched off by 48 hours and *POU5F1* by 72 hours, but this occurred to a significantly lesser extent on treatment with ligands and this was more pronounced with PhenDC3. This expression of pluripotency markers in the PhenDC3 treated samples was higher and indicative that these treated cells were not fully differentiating, or least not as efficiently as the DMSO treated cells.

The expression of EMT associated genes was monitored and compared between the samples (Figure 88b). The expression of E-cadherin (*CDH1*) is expected to decrease at the EMT and be replaced with N-cadherin (*CDH2*) (Teo et al., 2011). Both switching off of E-cadherin and turning on of N-cadherin occurred to a lower extent in the PhenDC3 treated cells which may explain why they did not completely differentiate, possibly having not completed the epithelial mesenchymal transition and causing the cells to remain in a mesenchymal state. The expression of *SNAI1*, *SNAI2* and *TWIST1* was also higher in the PhenDC3 treated cells, suggesting that the EMT was perturbed, there were some changes in the PDS treated cells but to a lower extent.

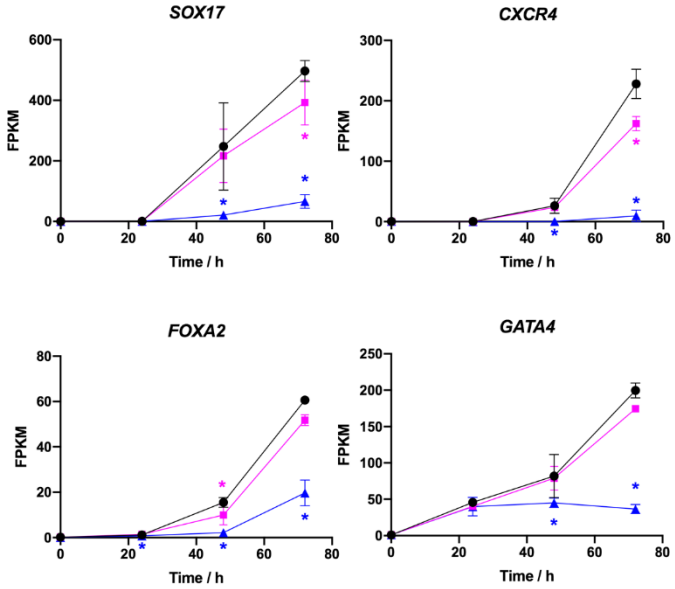
Expression of the proteins associated with mesendoderm, and therefore early differentiation markers were compared (Figure 88c). Expression of these genes had altered kinetics in both PDS and PhenDC3 treated cells. *GSC* was very clearly being expressed to a lower level in both conditions.

All definitive endoderm markers showed a clear trend: PhenDC3 treated cells did not express any of these marker genes as highly as the DMSO treated cells and PDS expressed slightly lower levels (Figure 88d). This suggested that SOX17 expression had been a good efficiency read-out for the endoderm differentiation and that both of the G4 ligands were perturbing this process, especially in the PhenDC3 treated cells.

Surprisingly, genes that had been used for qPCR normalisation and western blotting loading controls also changed during the differentiation process suggesting that these were a bad set to use (Figure 88e). Some of these were also altered with the ligands, further suggesting that permeabilised flow cytometry had been a better read-out for monitoring differentiation as discussed in Results II.


(a)
(b)
(c)
DMSO 2 µM
PDS 2 µM
PhenDC3 2 µM
POU5F1
SOX2
NANOG
CDH1
CDH2
TWIST1
SNAI1
SNAI2
EOMES
MIXL1
TBXT
FGF8
GSC
FPKM
Time / h

(d)



(e)

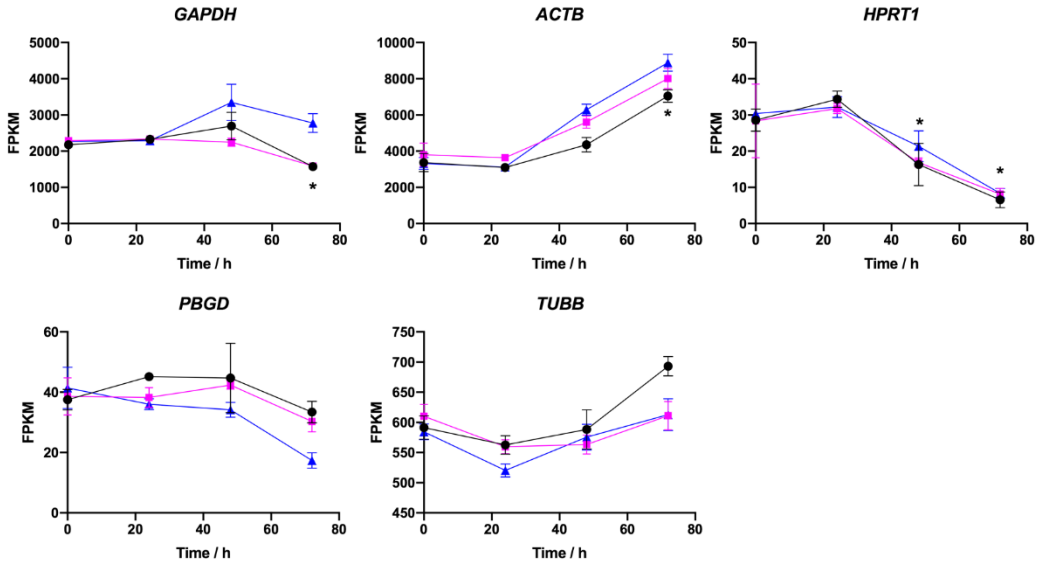


Figure 88. Normalised transcripts were calculated as fragments per kilobase of transcript per million mapped reads (FPKM) were calculated at each time point for each treatment condition and plotted. The mean is plotted, and error bars show the standard deviation of RNA triplicates. The triplicates were performed at the same time but using different biological samples. DMSO (to the equivalent volume of ligand) is shown in black, PDS in pink and PhenDC3 in blue. The zero-hour timepoint is taken to be the 24-hour undifferentiated sample, the rest are shown during differentiation. Significance is calculated for $p<0.05$ using the CuffDiff statistics test. Pink * (PDS) and blue * (PhenDC3) show significant difference compared to DMSO apart from in (e) where significance is DMSO treatment compared to the undifferentiated control (a) pluripotency genes (b) EMT-associated genes gene name for E-cadherin *CDH1*, gene name for N-cadherin *CDH2* (c) mesendoderm genes (d) definitive endoderm specification genes (e) qPCR and western blotting normalisation genes.

5.3.2.1 Enrichment analysis in G4 ligand treated cells during differentiation

The top ten GO terms and transcription factors were analysed for differentiation in the presence of both drugs at 72 hours and are shown below in Table 21. Most of the GO terms associated with the PDS treatment were developmental, showing that this treatment can alter differentiation. However, with PhenDC3 treatment the GO terms were associated with organelles and metabolism which is likely due to perturbing mitochondrially-encoded genes, for example a number of tRNAs and *COX* genes. The cells also showed altered metabolism from the medium colour and this may be due to remaining in the state of transferring through the EMT. These GO terms may also be due to the prolonged perturbed signalling pathways in this treatment. The overlap between these two conditions showed a large enrichment in developmental genes, suggesting that these are perturbing the same pathways and both alter the outcome of differentiation.

GO term	Description	Count	Adjusted p-value
PDS			
GO:0032502	Developmental process	117/6212	$2.01x10^{-14}$
GO:0048856	Anatomical structure development	170/5793	$2.01x10^{-14}$
GO:0007275	Multicellular organism development	160/5321	$2.01x10^{-14}$
GO:0032501	Multicellular organismal process	196/7414	$1.06x10^{-13}$
GO:0048513	Animal organ development	119/3428	$1.06x10^{-13}$
GO:0009653	Anatomical structure morphogenesis	97/2598	$2.44x10^{-12}$
GO:0048731	System development	142/4760	$6.76x10^{-12}$
GO:0008283	Cell proliferation	83/2121	$3.04x10^{-11}$
GO:0048646	Anatomical structure formation involved in morphogenesis	57/1144	$4.37x10^{-11}$
GO:0009887	Animal organ morphogenesis	51/969	$1.26x10^{-10}$
PhenDC3			
GO:0044424	Intracellular part	1883/14266	$9.05x10^{-37}$
GO:0005622	Intracellular	1906/14606	$5.97x10^{-34}$
GO:0043229	Intracellular organelle	1674/12497	$2.48x10^{-28}$
GO:0043231	Intracellular membrane-bounded organelle	1467/10736	$2.43x10^{-24}$
GO:0031323	Regulation of cellular metabolic process	938/6223	$1.54x10^{-23}$
GO:0043226	Organelle	1736/13288	$1.54x10^{-23}$
GO:0044237	Cellular metabolic process	1459/10713	$2.15x10^{-23}$
GO:0019222	Regulation of metabolic process	1009/6819	$2.32x10^{-23}$
GO:0080090	Regulation of primary metabolic process	924/6175	$5.64x10^{-22}$
GO:0005515	Protein binding	1540/11522	$6.57x10^{-22}$
Overlap of PDS and PhenDC3			
GO:0009653	Anatomical structure morphogenesis	57/2598	$1.14x10^{-8}$
GO:0001568	Blood vessel development	29/743	$1.14x10^{-8}$
GO:0048514	Blood vessel morphogenesis	28/664	$1.14x10^{-8}$
GO:0035295	Tube development	34/1079	$2.07x10^{-8}$
GO:0072358	Cardiovascular system development	29/788	$2.07x10^{-8}$
GO:0001944	Vasculature development	29/779	$2.07x10^{-8}$
GO:0023051	Regulation of signaling	66/3482	$2.07x10^{-8}$
GO:0010646	Regulation of cell communication	66/3447	$2.07x10^{-8}$
GO:0048856	Anatomical structure development	90/5793	$2.83x10^{-8}$
GO:0032502	Developmental process	93/6212	$5.48x10^{-8}$

Table 21. GO term enrichment of genes with deregulated expression at 72 hours in differentiation when treated with the ligands compared to the DMSO treated cells. The top panel shows the ten most significant GO terms upon PDS treatment, the middle on PhenDC3 treatment and the lower panel shows the enrichment of the overlap of the two treatments.

TRANSFAC TFBS	Factor	Motif	Count	Adjusted p-value
PDS				
TF:M01587	FPM315	SRGGGAGGAGGN	91/3242	0.000296
TF:M00986	Churchill	CGGGNN	277/14245	0.000296
TF:M02105	NF-E4	CHCCCTCKCCWG	76/2635	0.000654
TF:M04863	TF3C-beta	CCNGGAGGGCTTCCTGGAGGAG	186/8554	0.000654
TF:M01860	AP-4	NCAGCTGYNGNCN	142/6048	0.000654
PhenDC3				
TF:M00716	ZF5	GSGCGCGR	1964/14128	$4.39x10^{-71}$
TF:M04866	hdac2	CGCGCGCGC	2016/14708	$4.39x10^{-71}$
TF:M00803	E2F	GGCGSG	1856/13193	$8.01x10^{-63}$
TF:M04869	Egr-1	GCGCATGCG	1518/10244	$3.86x10^{-52}$
TF:M00695	ETF	GVGGMGG	1533/10416	$4.97x10^{-51}$

Table 22. The top five most enriched transcription factor binding sites in the genes deregulated on PDS (top) or PhenDC3 (bottom) treatment at 72 hours. There were no significantly enriched transcription factor binding sites in the overlap of the two treatments.

The most significant transcription factor binding sites in both treatments was analysed (Table 22). Interestingly, the transcription factor binding sites associated with PhenDC3 treatment gene deregulation were enriched in guanine bases. Although no statistical tests were performed, this observation suggests that PhenDC3 could be interfering with transcription factor binding sites and altering gene expression in this manner. Further analysis is being performed.

The transcripts which were deregulated compared to DMSO at every timepoint when treated with ligands are shown in the Venn diagrams (Figure 89a and b). In the PDS treated samples there were only eleven genes and the GO enrichment for these were all significantly associated with DNA damage and programmed cell death. Interestingly, some of these genes have found to be upregulated by p53 in unperturbed mESC differentiation, such as *CDKN1A*. In the PhenDC3 treated cells there were 26 deregulated transcripts shared at each timepoint. The significantly deregulated GO terms for this gene list were all associated with the respiratory chain, oxidative phosphorylation and the mitochondria. This suggests that the common deregulation at each time was associated with the ability of this ligand to alter the mitochondrial genome or alter signalling pathways involved in reprogramming metabolism. However, just comparing 48 and 72 hours shows much more highly significant GO terms associated with developmental process, suggesting that this ligand impacts differentiation in the latter stages.

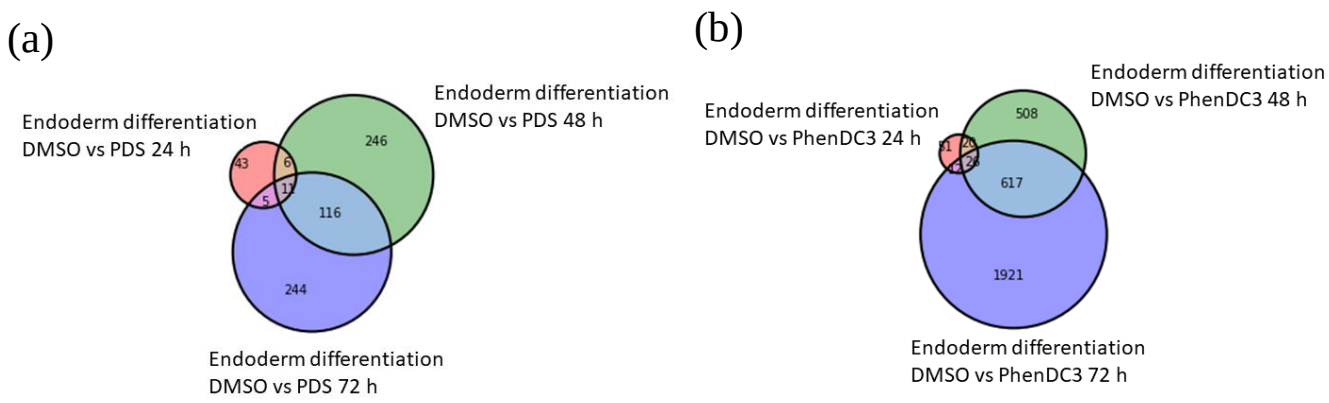
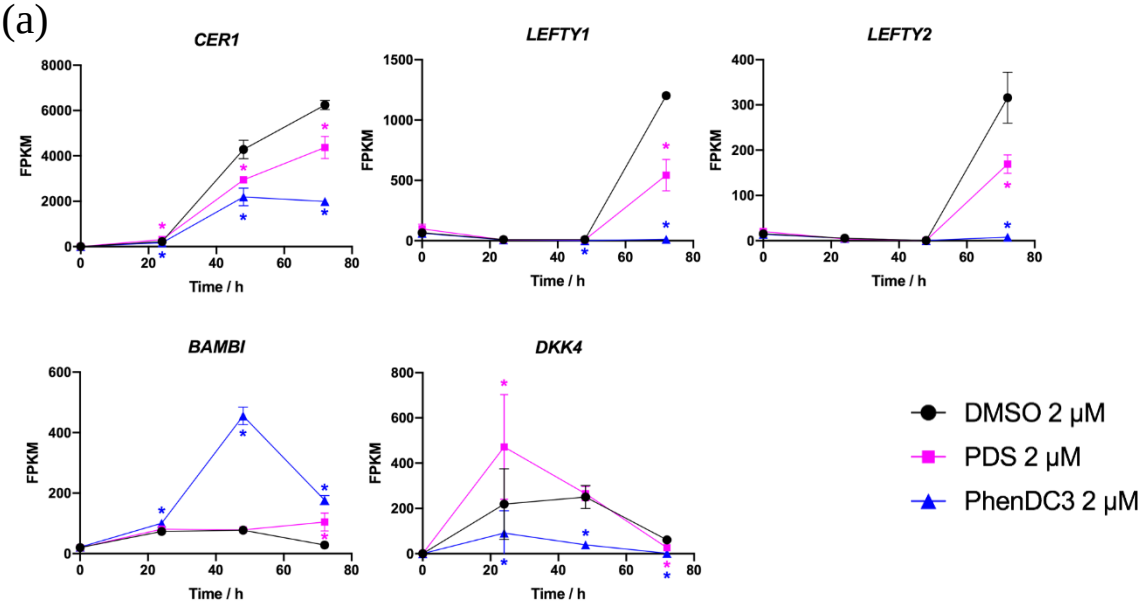


Figure 89. A comparison of the genes differentially expressed between DMSO and the G4 ligand treatment every 24 hours during differentiation (a) PDS treatment (b) PhenDC3 treatment.

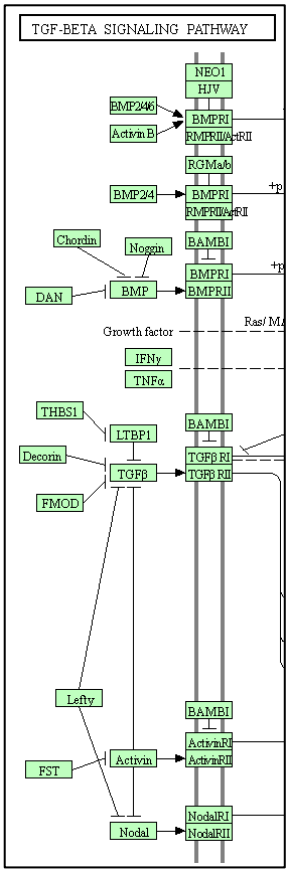
5.3.2.2 Key pathway changes induced by the G4 ligands during differentiation

The PhenDC3 treated cells expressed deregulated levels of WNT and TGFβ pathway transcripts throughout differentiation, compared to the DMSO treated samples, and PDS showed similar changes but to a lower extent (Figure 90a). The KEGG pathways are shown in (Figure 90b and c). The CHIR90021 GSK3 inhibitor was added during the first 24 hours of differentiation and this upregulates WNT pathway genes. The cells then produce WNT pathway inhibitors such as *CER1* and *DKK*. TGFβ pathway inhibitors such as *LEFTY1* and *LEFTY2* are produced to high levels during differentiation. However, it was clear that these inhibitors were turned on to a lower extent in the drug treated cells, and this was extremely pronounced in PhenDC3 treated cells. In the PhenDC3 treated cells, *BAMBI*, a WNT pathway activator was also upregulated, suggesting an enhancement in this pathway.

(a)



(b)



(c)

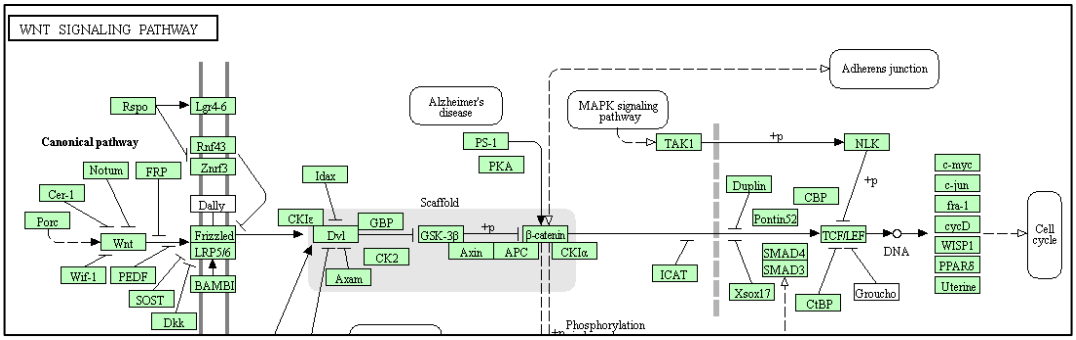


Figure 90. Analysis of deregulated pathways in ligand-treated differentiating cells (a) RNA sequencing reads of WNT and TGFβ signalling pathway transcripts during differentiation. The mean±SD are plotted and differentially expressed genes are marked as * ($p<0.05$ measured using CuffDiff differential expression pipeline) in comparison with the DMSO treatment at the same point in differentiation, PDS (pink) and PhenDC3 (blue) (b) the TGFβ signalling pathway taken from KEGG (c) the WNT pathway taken from KEGG.

In order to see whether the maintenance of the WNT pathway in differentiation was able to mimic the addition of PhenDC3 to wildtype cells, cells were differentiated either removing the inhibitor at 24-, 48- or 72-hours (Figure 91). Maintaining the GSK3 inhibitor for longer than 24 hours caused the cells to proliferate to a greater extent while also decreasing the expression of SOX17, as seen in PhenDC3 treated cells. Therefore, an extension of the time that the WNT pathway was activated decreased the expression of endoderm specification markers. This experiment, along with the analysis of WNT pathway genes using RNAseq during differentiation in ligand-treated cells, suggests that PhenDC3 may partly exert its effect through this pathway. However, further analysis would be required to draw a definitive conclusion. IWP2, a WNT inhibitor, acts at the level of activation of the pathway by inhibiting porcupine. Since this was earlier in the WNT pathway it was used instead of CHIR99021 to prevent the PhenDC3 effects. However, the solubility of this inhibitor was not high and in initial trials the volume of DMSO affected the cellular differentiation, this was not used further. However, the WNT pathway expression changes were thought to stem from BMP changes and it was clear that this pathway was highly perturbed.

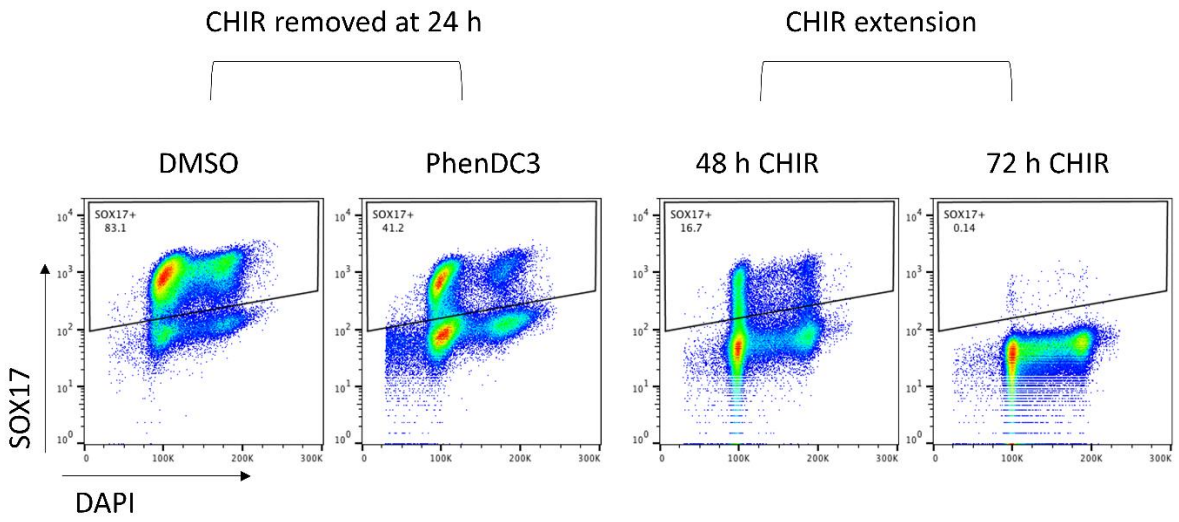


Figure 91. The WNT pathway activator CHIR99021 is added for the first 24 hours in unperturbed differentiation and then removed. To understand whether continual activation of the WNT pathway prevented efficient differentiation it was added for an extra 24 or 48 hours and the percentage of SOX17 expressing cells was monitoring at 72 hours using flow cytometry.

The EMT was clearly altered in the ligand treated cells. As the phosphorylation of H2A.X had remained up at 72 hours in the PhenDC3 treated cells, the level of H2AFX transcript in the RNAseq experiment was monitored (Figure 92). This data showed that the level was upregulated at 48 and 72 hours in the PhenDC3 treated cells but there is no direct evidence linking this to the EMT.

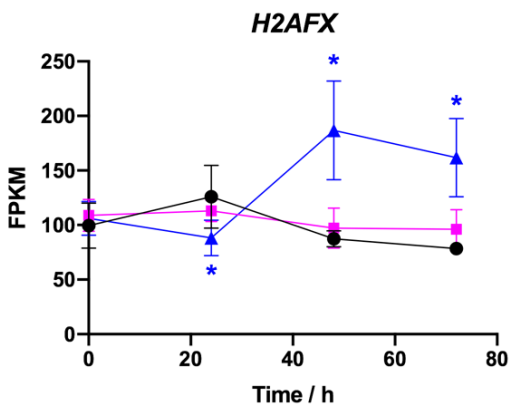


Figure 92. The level of *H2AFX* mRNA in ligand-treated and mock-treated cells during differentiation, taken from the RNA sequencing data. Mean±SD are plotted, with n=3 collected on the same day. Differential expression is calculated using the CuffDiff program in comparison to the DMSO treatment at each point in differentiation. Blue * denote significantly different expression of *H2AFX* on treatment with PhenDC3.

There were also a number of mitochondrially-encoded genes which have their transcripts deregulated upon ligand treatment, including *ATP8* and *ND4*, although a stress response did not seem to be induced. Table 21 shows an enrichment for metabolic processes and organelles, suggesting that there may be a primary effect of G4-stabilising on this genome. There seems to be an altered metabolism in these PhenDC3 treated cells but this may be due to changes in a number of signalling pathways, including in the EMT.

5.3.2.3 Lineage specification in G4 ligand treated cells

While it was clear that PhenDC3 cells were not differentiating as efficiently towards endoderm specification, and PDS to a lesser extent, gene expression changes were checked to analyse whether these treated cells were differentiating to a different lineage: mesoderm or neuroectoderm. This was not easy to analyse as there were no positive controls for the expression of these genes. Mesodermal gene expression changes were checked as the initial pathway specification is similar in both processes (Figure 93a). The level of *HAND1* was higher in both PhenDC3 and PDS treated cells at 48 and 72 hours. This change was more pronounced when looking at *IGFBP3* expression at 72 hours, where both PDS and PhenDC3 had very upregulated expression. Interestingly, both of these transcripts were significantly upregulated in the wildtype MMS treated cells at 72 hours in the RNAseq experiment from Results II. While these gene expression changes do not imitate the kinetics of differentiation to mesodermal lineage, this may suggest that when endoderm differentiation is perturbed the cells differentiate towards the mesodermal lineage.

The neuroectodermal lineage genes were also checked in order to see whether the cell fate was being specified in this direction (Figure 93b). As before, there were no positive controls so it was hard to know what the level of expression would be in a neuroectodermal cell. PDS may

have slightly increased the expression of *NES* but this was not the case in PhenDC3 treated cells. *SOX1* and *FGF5* were expressed at very low levels with all treatments. This suggested that the ligands were not altering the lineage specification to ectoderm, probably due to divergence of these pathways *in vitro*. Expression of 14 transcripts of each lineage were compared with each treatment (Figure 93c) which showed that definitive endoderm genes were expressed to the highest level in general, as expected from the differentiation protocol. Definitive endoderm genes looked to be expressed lower in both treatments, but this was more pronounced with PhenDC3. The mesoderm analysis showed that both PDS and PhenDC3 were upregulating these transcripts with respect to the DMSO treatment. No difference was seen in the ectoderm specification pathway, likely due to the culture conditions and the general low expression of these transcripts. The probability associated with these changes is hard to interpret due to the large differences in the expression of the transcripts in FPKM, especially as a value below one suggests no expression and many of the ectoderm genes are lowly expressed.

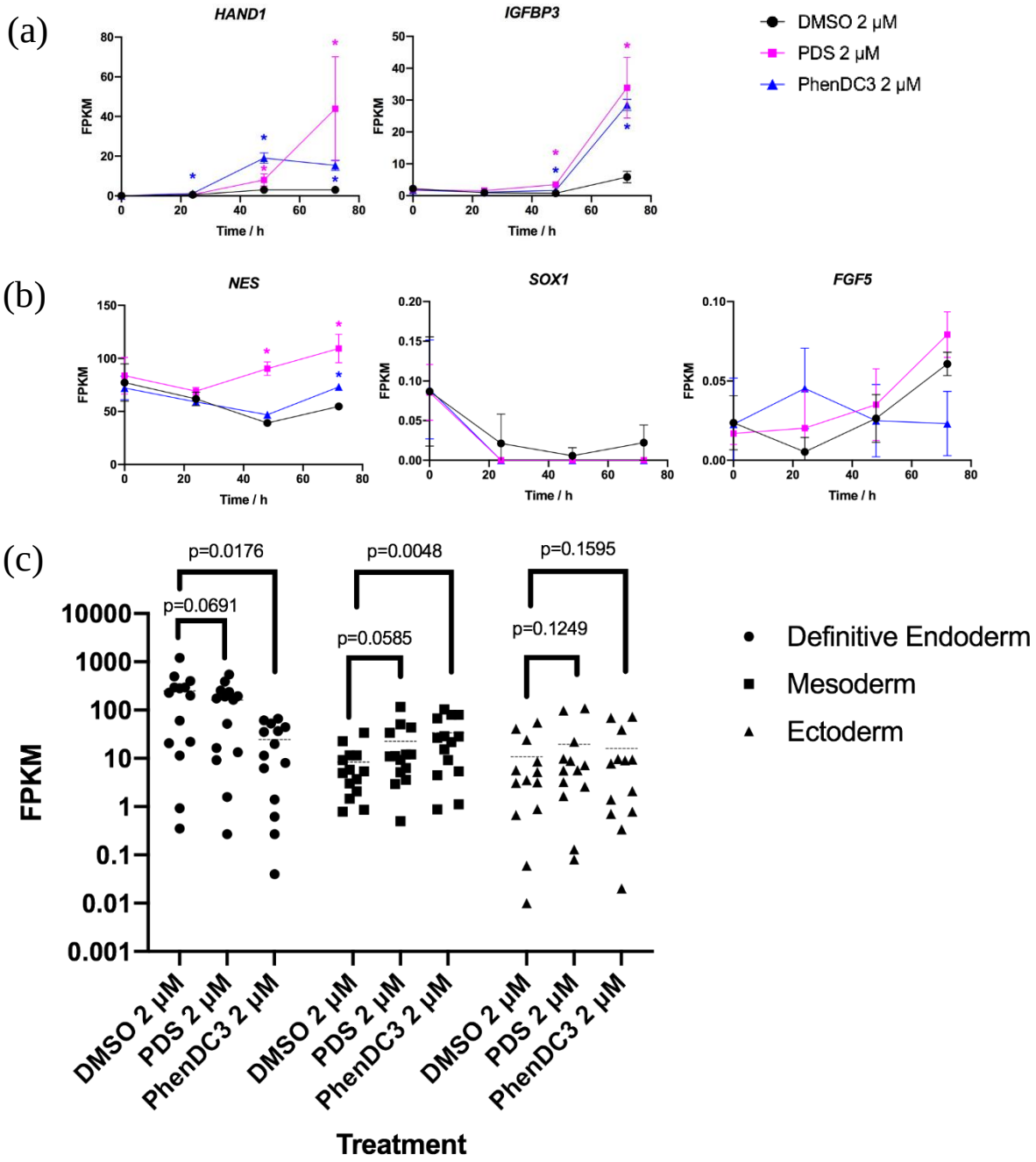


Figure 93. Analysis of genes involved in lineage pathway specification (a) mesoderm, mean±SD are plotted, differential expression is calculated using CuffDiff, $p<0.05$ compared to the DMSO treatment, PDS (pink) and PhenDC3 (blue) (b) as (a) but neuroectoderm genes (c) plots to show the expression of lineage specific marker genes at 72 hours in the different treatments. 14 transcripts known to be expressed in each lineage had expression calculated in FPKM and the average is plotted. Significance was calculated using a paired t test and p values are shown above the comparison in the figure.

5.3.3 Recombinant proteins give insights into key changes induced by G4 ligands

In order to definitively show that the G4 ligands were having a G4-specific effect on gene expression, G-quadruplexes playing a role in differentiation would need to be removed using gene editing. However, during differentiation there were a huge number of genes that were upregulated or downregulated. There were also a vast number of genes deregulated during differentiation in the RNAseq data when the cells were differentiated in the presence of G4 ligands. It would be a laborious if not impossible task to remove every potential G4 in every deregulated gene. While a hierarchical approach is much better to narrow down the genes that are altered earliest during differentiation, this still does not give a reasonable number of G4s to remove. Therefore, in the first instance, recombinant proteins were added to ligand treated cells to see if it was the lack of these proteins which was preventing differentiation. The recombinant proteins used were the products of genes predicted to have G4s (Figure 94).

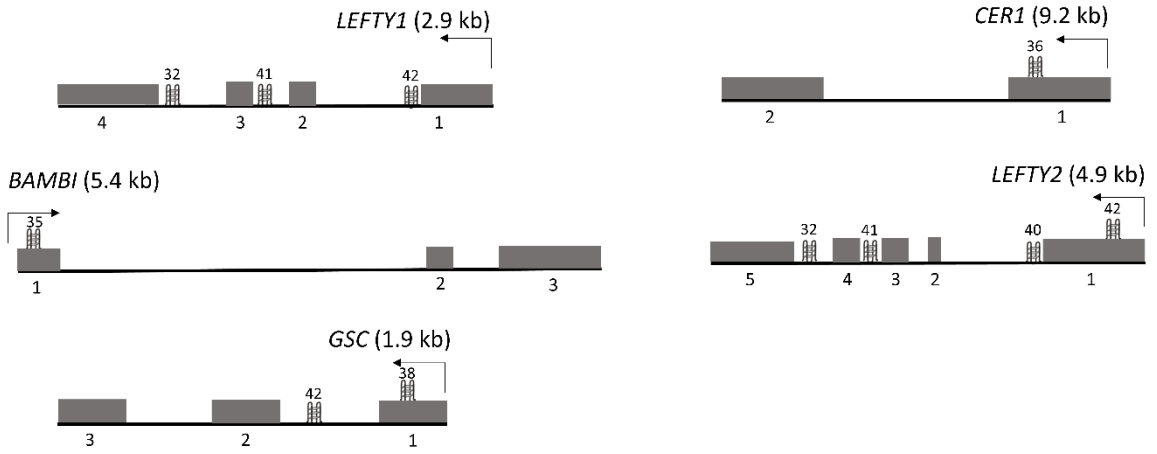


Figure 94. Mapping of potential G4s in genes shown to have altered expression on treatment with ligands. Predictions were performed using QGRS mapper using the motif $G_xN_{yi}G_xN_{y2}G_xN_{y3}G_x$ and using the assumptions that shorter loops are more common than larger loops, G4s tend to have loops roughly equal in size and the greater the number of guanines the more stable the G4. Predictions with a G-score ≥ 30 are shown, the greater the number, the higher the likelihood to form. Exons are not to scale, the length of the gene is marked above the TSS.

In order to check whether reconstituted proteins would be a good way forward, a media switch experiment was performed in a similar manner to that described by Kempf et al, in the Nature Communications paper (Kempf et al., 2016). They showed that the secreted proteins produced by cells in the first six hours were crucial in directing differentiation. This experiment was performed such that untreated cells were differentiated for six hours and then this media had DMSO, PhenDC3 or PDS added and was transferred to a new well to set up differentiation (Figure 95). This experiment showed that the PhenDC3 treated cells expressed a much higher level of SOX17 when the media had been switched, in contrast to DMSO and PDS. This

suggested that PhenDC3 was preventing an early expressed secreted protein from being expressed and its expression was able to bypass some of the problems occurring in differentiation with PhenDC3. This was in line with the results that showed PhenDC3 needed to be present in the first 24 hours of differentiation in order to induce a change to differentiation. However, the fact that the first six hours were important suggested that this may not be a G4-replication dependent, epigenetic phenomenon as the timescale was too short. The paper also highlighted CER1 and LEFTYA as key secreted components and since both were not upregulated in the PhenDC3 treated cells, these recombinant proteins were used.

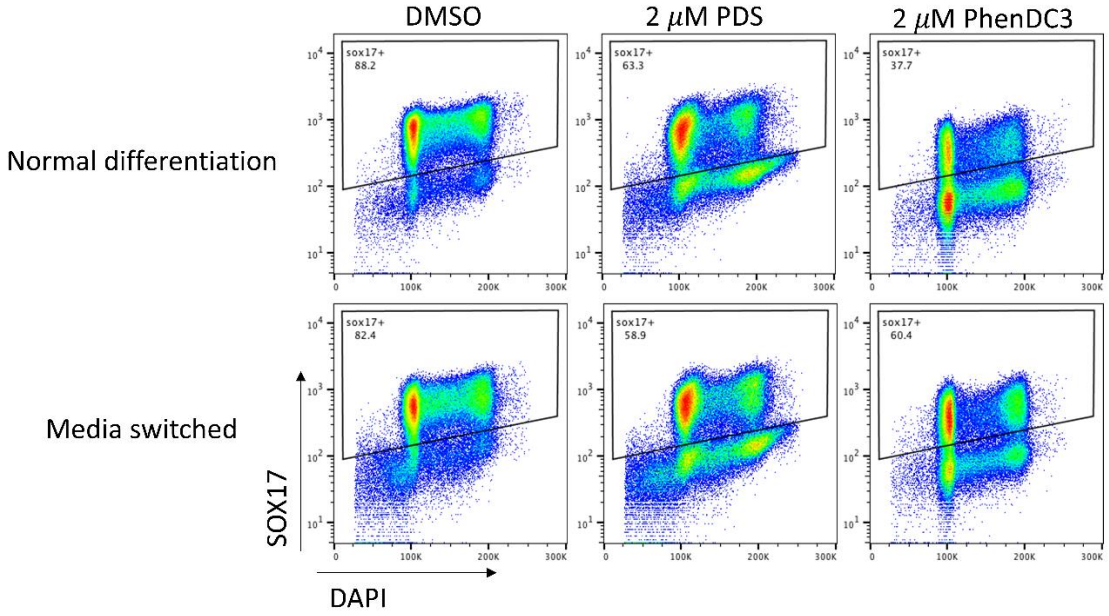


Figure 95. Media switching differentiation experiment to show the importance of secreted factors on the outcome of differentiation. Cells were differentiated for six hours in the presence of DMSO and then this media was added to undifferentiated cells to initiate differentiation in these cells, DMSO, PDS or PhenDC3 were also added when the media was switched, and the cells were differentiated for a further 72 hours to assess the efficiency of differentiation. Efficiency was analysed by the proportion of SOX17 expressing cells at 72 hours.

CER1 gene expression did not increase in the PhenDC3 treated cells as in the DMSO treated cells during differentiation (Figure 90a). The cerberus protein is known to act as an inhibitor of the WNT pathway, acting during embryonic differentiation. Since this transcript was not upregulated in the PhenDC3 treated cells, the recombinant protein CER1 (R&D) was added during the differentiation process to see if this could compensate, at least partly, and allow cells treated with PhenDC3 to differentiate more efficiently. The CER1 protein was added to the cell culture medium (100 or 200 ng/mL) from 24 to 72 hours, with and without PhenDC3 (Figure 96). LEFTYA (R&D) recombinant protein was also added individually and in combination with CER1 for the final 48 hours of differentiation (Figure 96). There was a small increase in SOX17 protein expression with the addition of both CER1 and LEFTYA in the PhenDC3 treated cells,

but this was not enough to rescue the 25% loss in the treated cells and the increases were only small. The combination of the two recombinant proteins did not increase the differentiation efficiency which was surprising. This did not show convincingly that reconstitution of these proteins rescued the experiment, however despite the fact that it was shown to be upregulated at 48 hours in the RNA sequencing data, the expression may have been needed from six hours or earlier if this is in line with the media switch experiments. This experiment needs to be repeated, adding CER1 and LEFTYA from the onset of differentiation to understand whether a lack of these proteins is contributing to the PhenDC3 treated cell phenotype.

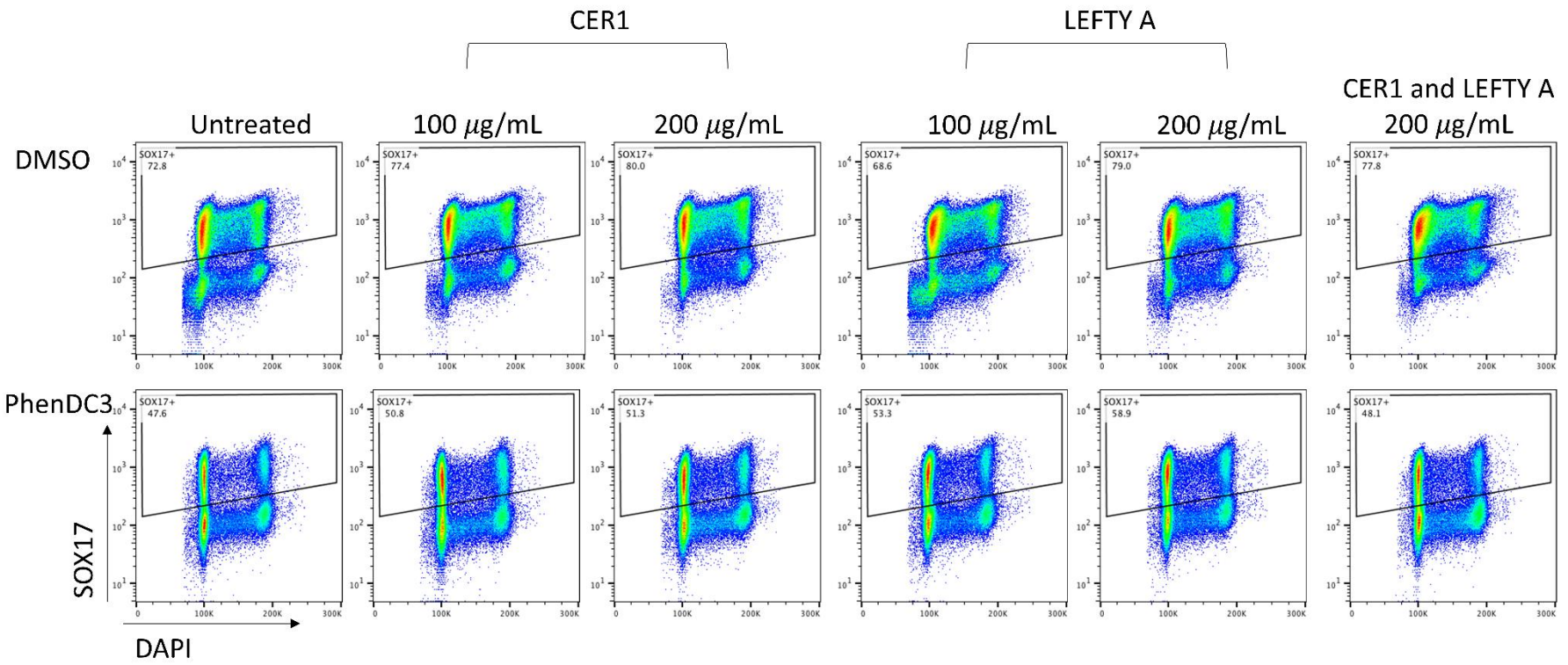


Figure 96. The addition of recombinant proteins during differentiation on treatment with PhenDC3 and DMSO in order to understand whether this could counteract the addition of PhenDC3. CER1 (left), LEFTYA (middle) and a combination (right). CER1 was added at 24 hours and LEFTYA at 48 hours. SOX17 expression was checked at 72 hours.

5.4 Initial differentiation of hESCs in the presence of G4 ligands

The H9 hES cells were previously verified to differentiate efficiently, Section 4.5, and they responded to MMS in a similar although not identical manner to the BOBSC cells. In order to compare the response to G4 ligands in differentiation, these experiments were repeated in the presence of PDS and PhenDC3.

5.4.1 Differentiating H9 hESCs with PDS and PhenDC3

H9 cells were differentiated in the presence of PDS and compared to the hiPS cells (Figure 97a). The cells appeared to react to the ligand in a similar manner, there was increased cell death in the differentiating wells compared to the undifferentiated. The percentage of SOX17 expressing cells also decreased, although not to as great an extent as the BOBSC cells. However, the PDS treated H9 cells not expressing SOX17 did not have an increased proportion of G2/M phase cells, unlike the BOBSC cell line. This suggested that the cells did not alter the cell cycle in the same way in the response to treatment.

A dose response of endoderm differentiation in BOBSC and H9 cells to PDS and PhenDC3 was performed to see if the cells tolerated different concentrations of ligands (Figure 97b). By eye the H9 cells reacted in a similar manner as BOBSCs to the ligands: PDS treated cells had increased cell death and cell size whereas PhenDC3 treated cells had a higher metabolism seen from the colour of the medium and the number of cells. This shows that while both the cell lines did exhibit a dose response to PDS and PhenDC3, the H9 cells were able to tolerate much higher doses of both G4 ligands and therefore the response was different. The reason behind these differences would be interesting to study and may be related to the differences in gene expression in the differentiation processes in normal differentiation, shown by the increased upregulation of EOMES discussed in Section 4.5.1.

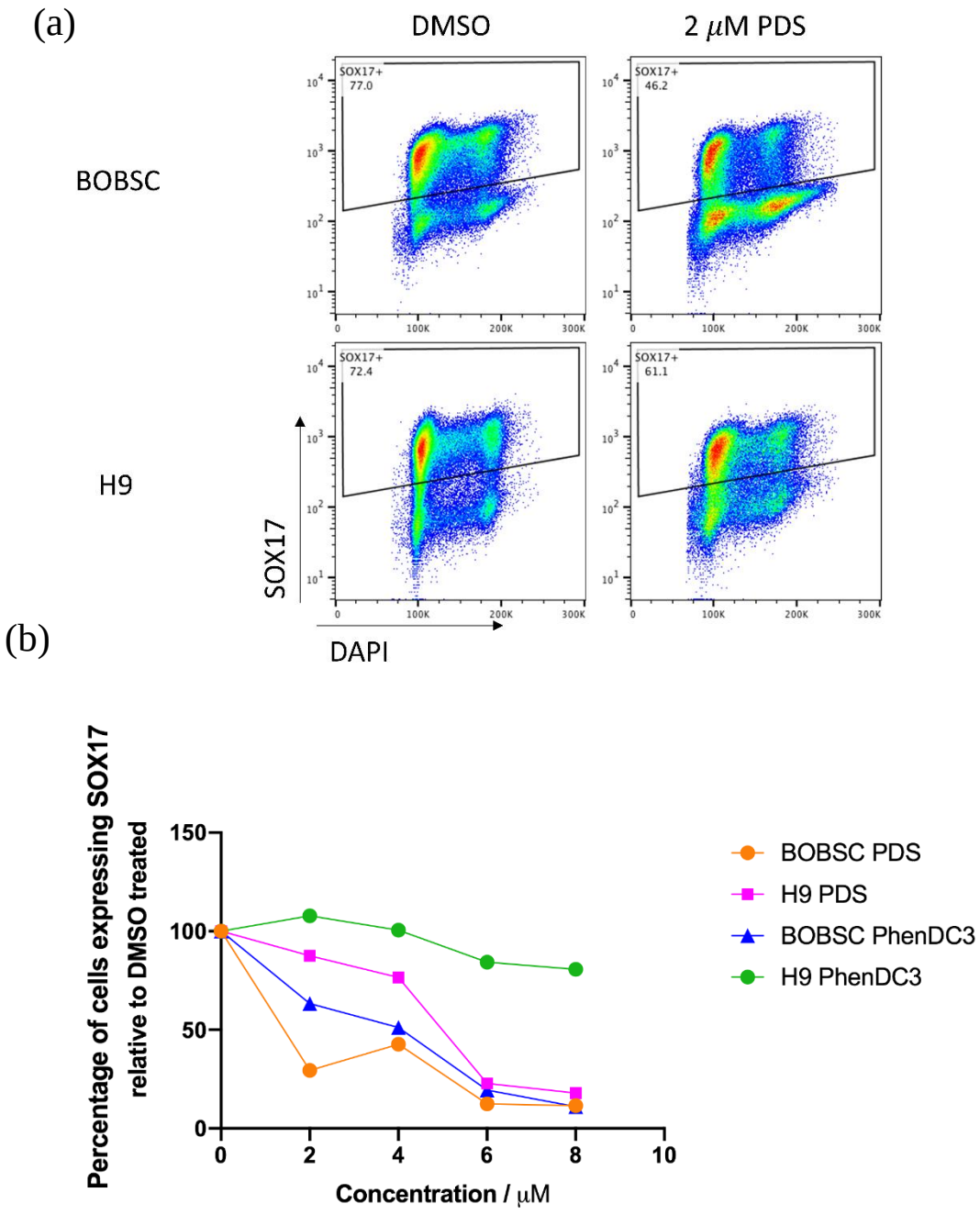


Figure 97. Analysis of G4 ligand treatment of H9 hESCs during differentiation (a) differentiation in the presence of 2 μM PDS and analysis of the proportion of SOX17 positive cells at 72 hours (b) the percentage of SOX17 positive cells relative to mock-treated cells at 72 hours in H9 and BOBSC cell lines treated with 2 μM PDS or PhenDC3, one experiment.

Chapter 6. Results IV: Genetic knockouts of proteins involved in the processing of G-quadruplexes

6.1 The G4-processing enzyme knockout BOBSC cell lines

I used the knockout BOBSC cell lines of enzymes known to be involved in processing G4s generated by the Sanger Centre: *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$. The knockouts were verified using qPCR, with primers designed downstream of the targeted exon (Figure 98). These results showed that the transcript levels were much lower in the knockout cells and these cell lines were used throughout this chapter.

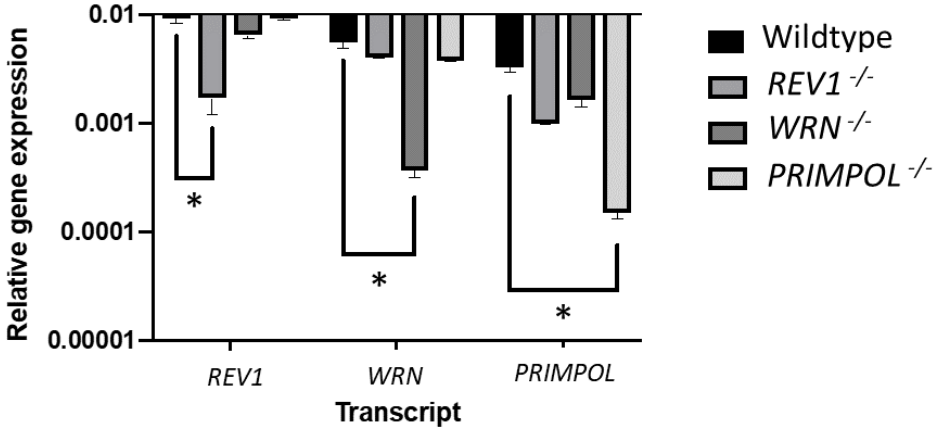


Figure 98. Verification of knockout cell lines from the Sanger Centre COMSIG project for *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ qPCR data normalised to *GAPDH*, bar orientation reversed for clarity. Mean and standard deviation of one biological replicate are shown, n=3 technical replicates, multiple t testing was used and * denotes $p<0.05$.

6.1.1 The undifferentiated state

6.1.1.1 The cell cycle of *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cells

Undifferentiated cell cycle profiles in each of these cell lines were performed, using the EdU-2D cell cycle plot, and compared to wildtype cells (Figure 99a). This approach showed that the *REV1*$^{-/-}$ cells had a slightly increased proportion of cells in G2/M phase of the cell cycle, mimicking the wildtype cells treated with PDS or DNA damaging agent (Figure 99b).

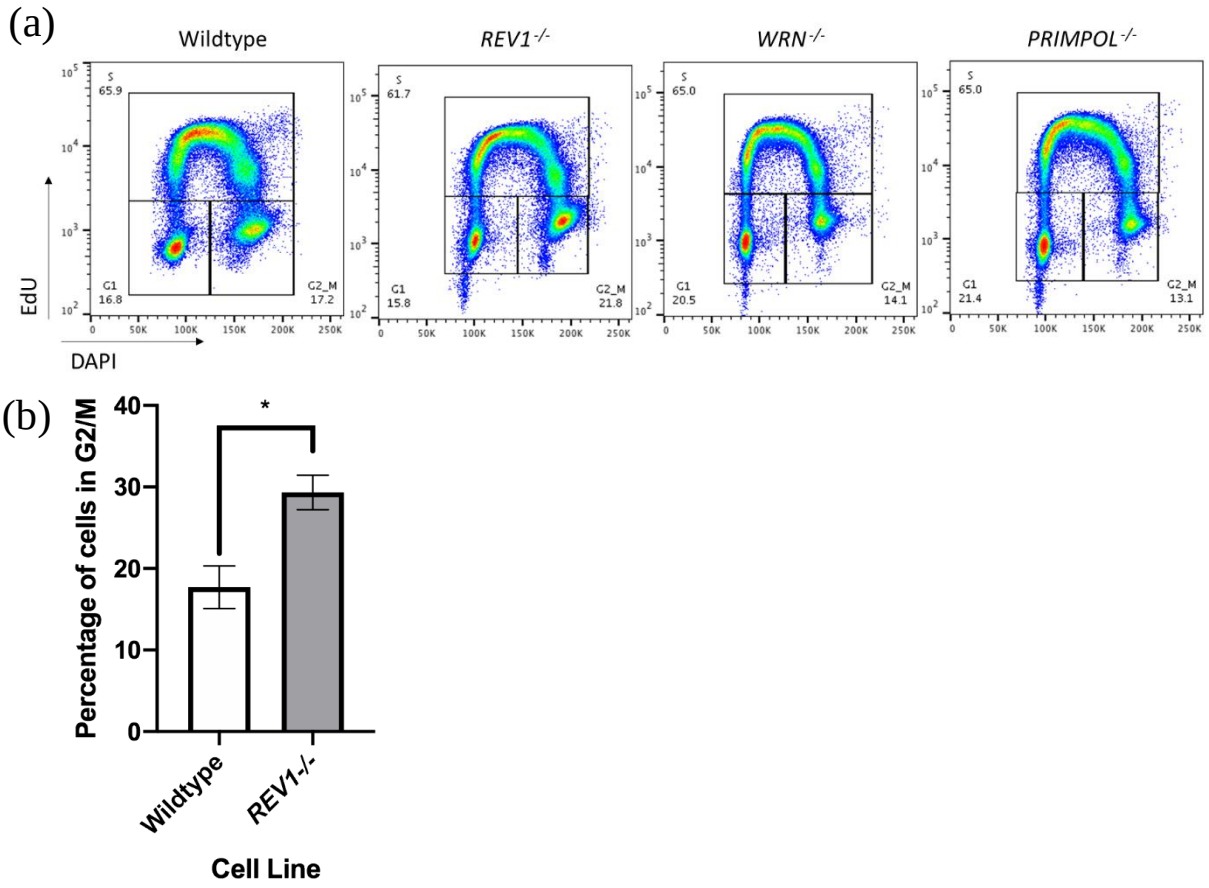


Figure 99. (a) 2D-cell cycle plots of EdU verses DAPI in genetic knockout cell lines in the undifferentiated state. *REV1*$^{-/-}$ shows an increase in the proportion of cells in G2/M of the cell cycle. (b) Quantification of the proportion of cells in G2/M phase of the cell cycle in *REV1*$^{-/-}$ cells compared to the wildtype cell line. Unpaired t test, p=0.0256, n=3, mean±SEM are plotted.

6.1.1.2 DNA damage response markers in the undifferentiated state

γH2A.X flow cytometry was performed on undifferentiated knockout cells and compared to the wildtype cell line (Figure 100a). Each of these cell lines showed a similar level of γH2A.X to the wildtype, suggesting that there was no accumulation of damage in these cell lines when cultured in the undifferentiated state.

Another DDR marker was checked using western blotting in undifferentiated *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cell lines: pCHK1 and CHK1 were analysed and compared with wildtype cells and wildtype cells treated with 2 mM HU to induce strong CHK1 phosphorylation (Figure 100b). Only the positive control cells gave a pCHK1 signal on the blot, suggesting that CHK1 is not phosphorylated in any of these three cell lines. Again, suggesting there is no base line damage in these cell lines. Surprisingly, the level of CHK1 in *REV1*$^{-/-}$ cells was lower than in the other cell lines. While it has been shown in Xenopus that REV1 is important for the ATR/CHK1 checkpoint (DeStephanis et al., 2015), there is no obvious reason why there would be a lower level of protein in this cell line. Further experiments would be required to understand whether this was a repeatable difference.

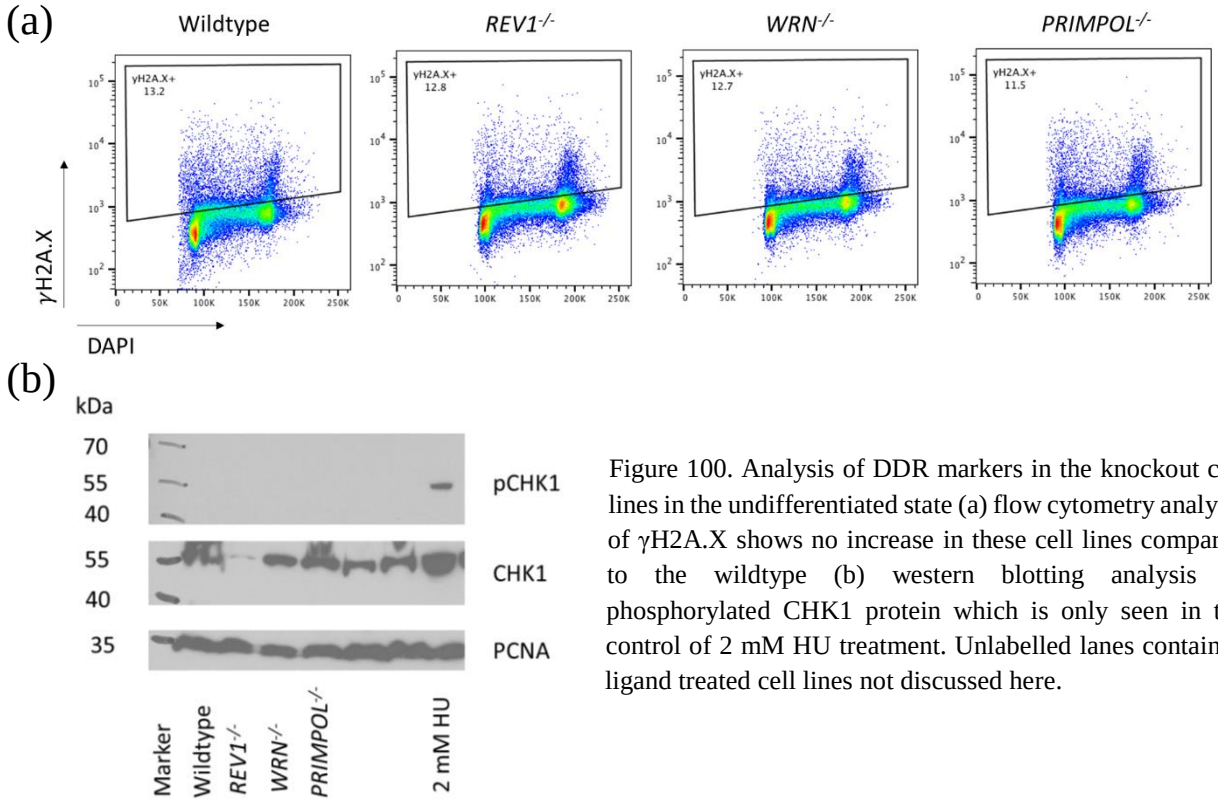


Figure 100. Analysis of DDR markers in the knockout cell lines in the undifferentiated state (a) flow cytometry analysis of γH2A.X shows no increase in these cell lines compared to the wildtype (b) western blotting analysis of phosphorylated CHK1 protein which is only seen in the control of 2 mM HU treatment. Unlabelled lanes contained ligand treated cell lines not discussed here.

6.1.1.3 Loss of adherence in the undifferentiated state

The knockout cells were cultured in the undifferentiated state to see if their epigenetic status, and gene expression profile, were altered over time. However, it was clear that there were increased numbers of *REV1*$^{-/-}$ cells that had lost adherence in the dishes, in a similar fashion to the ligand treated mESCs (Figure 101). While this loss of adherence is a very non-specific phenotype, this could suggest that *REV1*$^{-/-}$ cells were not maintained in the undifferentiated state as well as the wildtype cell line. If his were the case, a change in epigenetic state or gene

expression may have caused cells to lose adherence to the tissue culture dish which will be interesting to explore further.

	Viable cells / mL x 10^6	Viability / %	Total cells / mL x 10^6
Wildtype	0.267	29.6	0.903
REV1$^{-/-}$	0.485	32.6	1.488
WRN$^{-/-}$	0.234	31.0	0.757
PRIMPOL$^{-/-}$	0.214	36.2	0.592

	Viable cells / mL x 10^6	Viability / %	Total cells / mL x 10^6
Wildtype	1.29	32.8	3.92
REV1$^{-/-}$	5.39	31.2	17.29
WRN$^{-/-}$	2.85	29.5	9.65
PRIMPOL$^{-/-}$	1.31	31.3	4.18

Figure 101. A table to show two separate experiments whereby the cells which has lost adherence in each cell line were counted and the viability checked. The experiments were performed with different numbers of cells but in the *REV1*$^{-/-}$ cell line (red) there were greater numbers of total cells and total viable cells in this population, compared to all other cell lines.

6.1.1.4 Gene expression profiles of *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cell lines

These *rev1* and *primpol* DT40 knockout cell lines are known to have altered gene expression profiles owing to replication impediment induced epigenetic changes: BU1 protein loss occurs due to a G4 at the BU-1 locus. However, BLM and WRN are thought to be redundant in the DT40 system with respect to processing of G4s, such that there is only a defect in these double knockout cell lines (Sarkies et al., 2012). It would be very interesting if a similar pattern of gene expression changes occurred during maintenance of the undifferentiated state in these hiPS cells. In order to analyse the gene expression changes in these cells, RNA sequencing was performed. This set of sequencing was the first set performed in the thesis and therefore some interesting observations were learnt for further experiments. This data is discussed in Section 6.3.

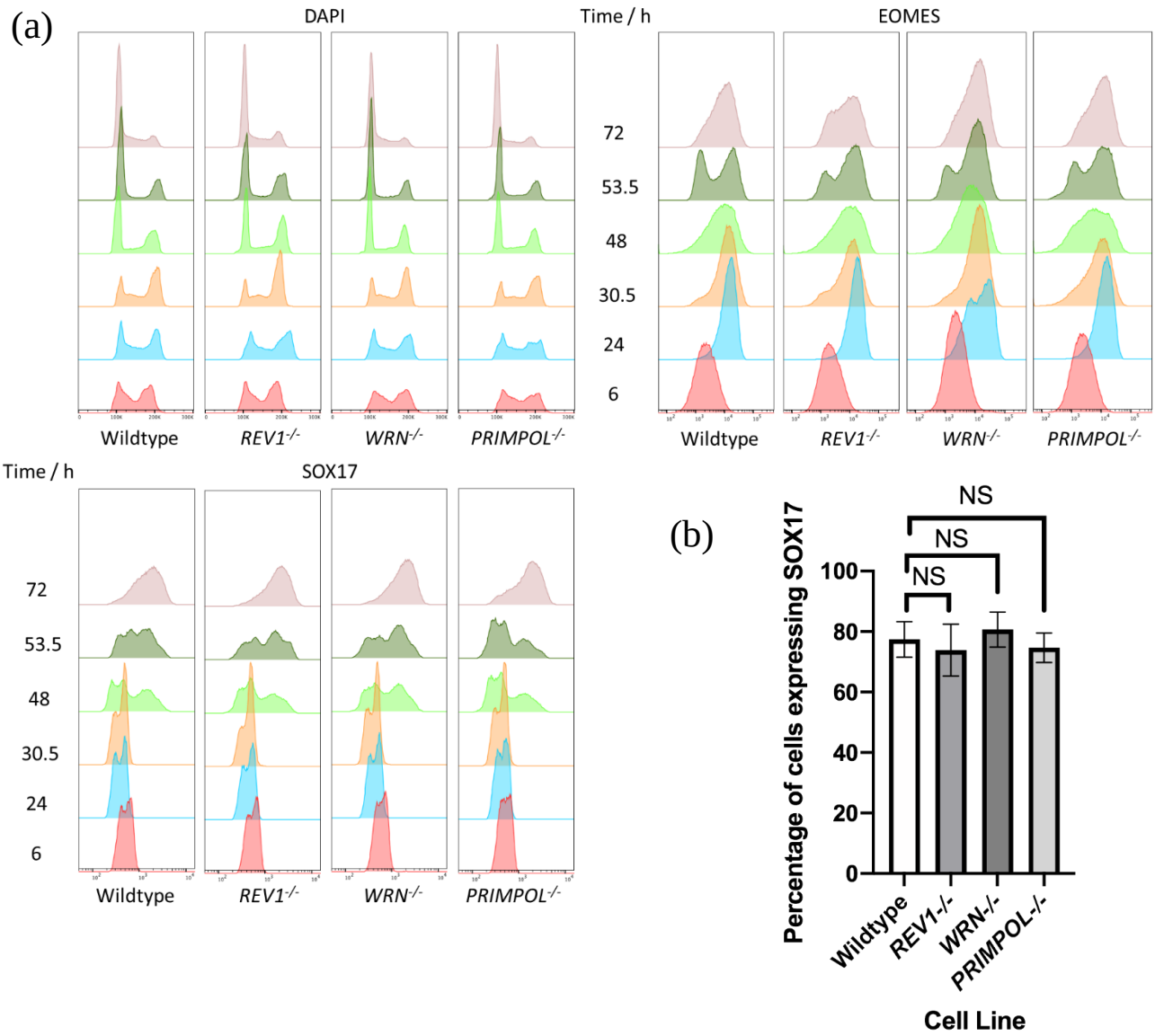


Figure 102. (a) Flow cytometry analysis of the differentiation to definitive endoderm from six to 72 hours of wildtype, *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cell lines. DAPI monitored DNA content, *REV1*$^{-/-}$ cells showed an increase in G2/M at 30.5 h (48.0 vs 40.2%). All cell lines had an increased population in G1 phase at 72 hours. EOMES was upregulated by 24 hours in all cell lines, SOX17 was expressed in all cell lines at 72 hours. (b) Percentage of cells expressing SOX17 at 72 hours, n=4, mean±SEM are plotted. Significance measured using unpaired t test, all changes were not significant compared to the wildtype cell line.

6.2 Definitive endoderm differentiation of knockout BOBSC cell lines

6.2.1 Differentiation of knockout cell lines

Each of the cells lines, *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$, were differentiated alongside wildtype cells and the differentiation was analysed every few hours using permeabilised flow cytometry (Figure 102a). The flow cytometry shows that each of the knockout cell lines differentiated efficiently: they expressed EOMES at 24 hours and SOX17 at 72 hours to high levels equivalent

to the wildtype (Figure 102b). This was in contrast to the cells differentiated in the presence of G4 ligands. This suggested that if these enzymes are required during the differentiation process, they are likely to be redundant. However, as seen previously, the cell cycle profile seen with DAPI staining showed that *REV1*$^{-/-}$ cells exhibited an increased proportion of cells in G2/M phase, especially around 30 hours into differentiation. This raised the possibility that these cells cause some form of replication impediment, in a similar manner to the PDS treated cells, and this may play a role during prolonged cell culture.

6.2.2 The γH2A.X spike in knockout cell lines

Wildtype, *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cells were differentiated and collected for permeabilised flow cytometry (Figure 103). This figure shows that each of the mutants showed an increased γH2A.X signal at 53.5 hours compared to the wildtype cells. This was three hours after the strongest upregulation in wildtype cells, suggesting that these cells may have gone through the EMT later than the wildtype, or were remaining at this transition for longer. However, since this was only monitored at 53.5 hours for this one experiment, the reproducibility cannot be commented on as the significance cannot be ascertained.

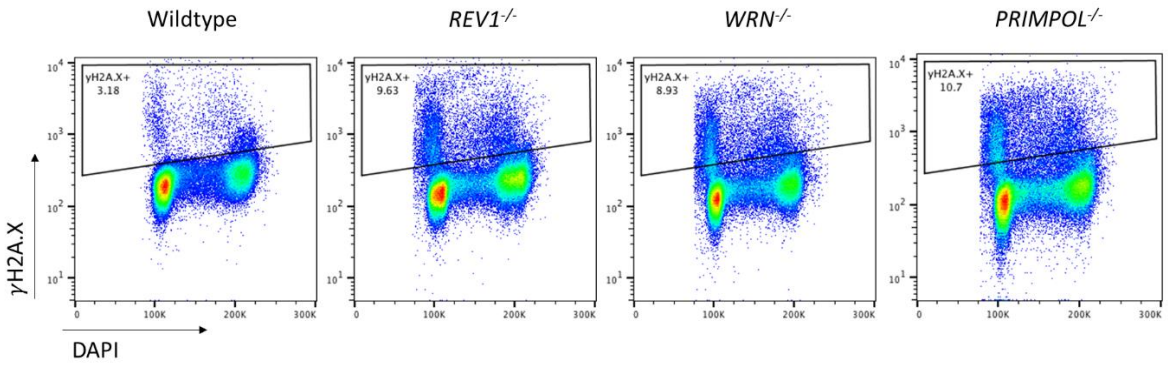


Figure 103. Flow cytometry plots to show the phosphorylation of H2A.X at 53.5 hours in the wildtype, *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cell lines during differentiation. DAPI staining on the x-axis labels the DNA context, the y-axis shows H2A.X phosphorylation. At this point each of the knockout cell lines showed higher levels of this marker.

6.3 RNA sequencing data reveals changes to gene expression profiles in knockout cell lines compared to the wildtype

In order to give a detailed understanding of the global transcriptome in these knockout cell lines, RNA sequencing was performed in the undifferentiated and differentiated state of *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$, as performed in Results II and III.

6.3.1 The undifferentiated state

This experiment was the first RNA sequencing experiment performed, therefore some of the information learnt was used in Results II and III when designing these sequencing experiments. Cells in the undifferentiated state were cultured for one month and RNA was extracted from each of the three knockout cell lines, and the wildtype cell line, at three separate timepoints. The theory behind this was that each replicate for a given cell line would be collected at a variety of points whereby the cells had different levels of confluence, were collected on different days of the week and at different times of day: giving the greatest variability in the results. RNA libraries were produced using RNA collected in this manner and sent to be sequenced as explained previously. The principal component analysis plot was produced and the result was surprising (Figure 104). This result showed that the 'batch effect' with these cells was larger than the gene expression difference between each of the knockout cell lines. This was likely to be due to the difference in confluence in these adherent cells and that there were not large differences in gene expression profile between these cells compared to the wildtype.

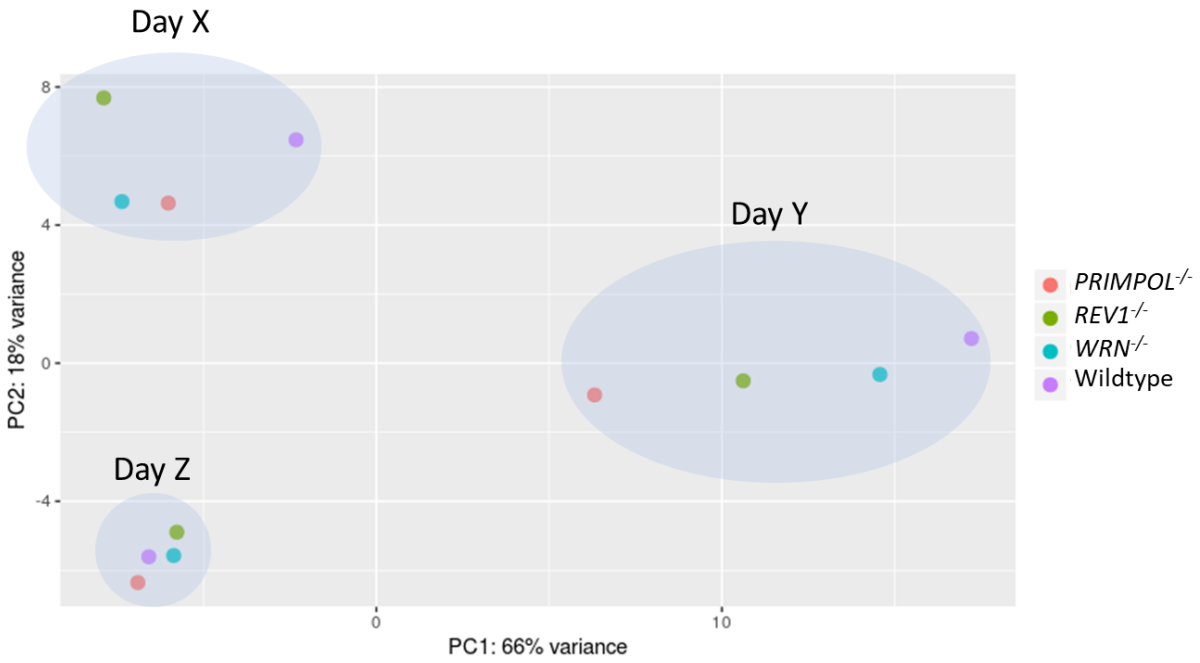


Figure 104. A PCA shows the variation within this dataset. Wildtype, *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cell lines were collected in the undifferentiated state on different days. These results show that the samples cluster together in terms of their expression in terms of the day collected, rather than with the different cell lines. This suggests that there was not much variation but in order to see variation the cells had to be collected on the same day.

However, to remove this effect, the experiment was repeated using RNA triplicates collected on the same day, at the same time, in all cell lines: this was used where possible in all sequencing experiments. In the undifferentiated state, there were not many deregulated genes compared to the wildtype cells: *REV1*$^{-/-}$ had 41, *WRN*$^{-/-}$ 17 and *PRIMPOL*$^{-/-}$ 58, summarised in Figure 105

and Table 23. This showed that there were very few overlaps between each of these cell lines which was not very surprising as there were not many deregulated genes in total.

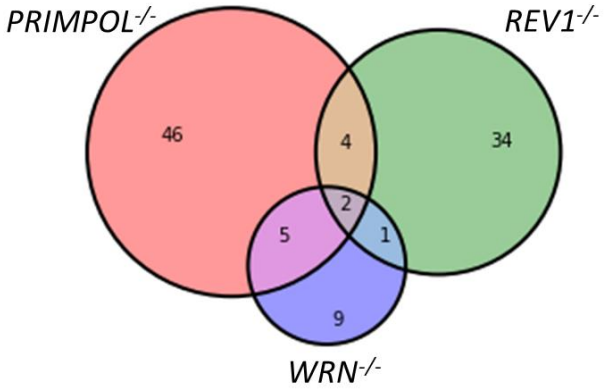


Figure 105. A Venn diagram to show the deregulated transcripts in the undifferentiated state in each cell line: *PRIMPOL*$^{-/-}$ in red, *REV1*$^{-/-}$ in green and *WRN*$^{-/-}$ in blue. The overlaps between each cell line are shown.

***REV1*$^{-/-}$ vs *PRIMPOL*$^{-/-}$**	***REV1*$^{-/-}$ vs *WRN*$^{-/-}$**	***PRIMPOL*$^{-/-}$ vs *WRN*$^{-/-}$**	***REV1*$^{-/-}$ vs *PRIMPOL*$^{-/-}$ vs *WRN*$^{-/-}$**
ESRP2 *C7orf43* *ENSG00000255026* *ENSG00000233056* *RNR1* *TEX41*	*C7orf43* *ENSG00000248215* *ESRP2*	*ESRP2* *C7orf43* *MAGI2* *CER1* *SCN8A* *PILRB* *CHCHD2*	*C7orf43* *ESRP2*

Table 23. A list of genes commonly deregulated between cell lines in the undifferentiated state. *C7orf43* is an uncharacterised open reading frame containing c-Myc and Sp1 transcription factor binding sites in the promoter. ESRP2 is epithelial splicing regulatory protein 2 which regulates splicing during the EMT.

The PCA of the RNA collected on the same day in all samples was performed and is shown in Figure 106. In this instance the triplicates clustered together well in the PCA and each genetic knockout cell line had an individual location on the PCA related to the gene expression state. This showed that the *PRIMPOL*$^{-/-}$ and *TP53*$^{-/-}$ cell lines grouped the furthest from the wildtype cell line, and also furthest from each other. This experiment showed that there were differences in the gene expression between these cell lines, but that the changes in confluence are great enough to override these differences.

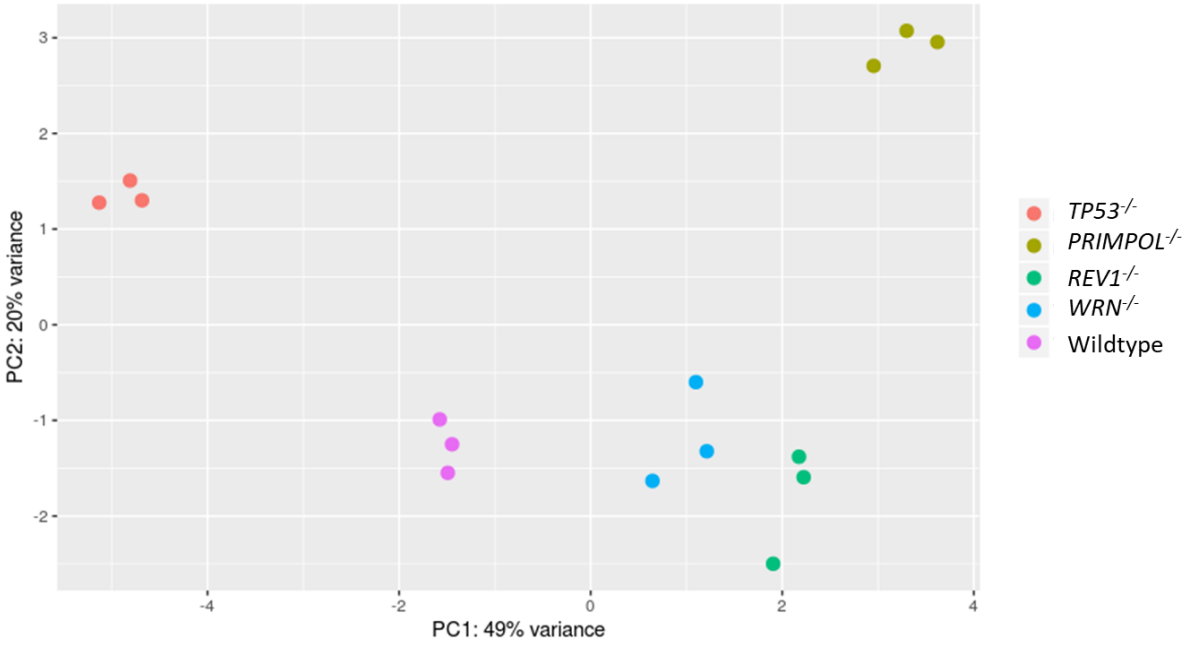


Figure 106. The PCA of undifferentiated knockout cell lines, as in Figure 104 but with RNA collected on the same day. *TP53*$^{-/-}$ was included for comparison as it was being performed for Results II. This showed that each cell line clustered together on the plot but the *PRIMPOL*$^{-/-}$ cell line had the greatest differences compared to the other cell lines.

The gene ontology analysis is described below in Table 24, with all enriched hits with a p-value ≤0.05 shown for each or the top ten, whichever had fewer values. The *WRN*$^{-/-}$ cell line had very few GO terms enriched, due to the lack of deregulated genes. The GO analysis associated with *REV1*$^{-/-}$ showed enrichment of nucleic acid processes, as expected from a polymerase. However, the GO enrichment for *PRIMPOL*$^{-/-}$ cells showed very significant p-values, many orders of magnitude lower than the other two cell lines. Most of these terms are associated with DNA and chromatin, which does appear to be promising. However, a number of histone proteins have deregulated expression levels in *PRIMPOL*$^{-/-}$ cells and it is known that these alter vast numbers of GO terms. Therefore, this is likely to be overrepresented by the histone proteins.

GO ID	Description	Count	Adjusted p-value
REV1$^{-/-}$			
GO:1902679	Negative regulation of RNA biosynthetic process	9/1221	0.0274
GO:1903507	Negative regulation of nucleic acid-templated transcription	9/1219	0.0274
GO:0045892	Negative regulation of transcription, DNA-templated	9/1171	0.0274
GO:0051253	Negative regulation of RNA metabolic process	9/1302	0.0339
GO:2000113	Negative regulation of cellular macromolecule biosynthetic process	9/1408	0.0421
GO:0045934	Negative regulation of nucleobase-containing compound metabolic process	9/1423	0.0421
GO:0031643	Positive regulation of myelination	2/13	0.0421
GO:0010558	Negative regulation of macromolecule biosynthetic process	9/1408	0.0455
WRN$^{-/-}$			
GO:0010225	Response to UV-C	2/14	0.0325
GO:0002467	Germinal center formation	2/15	0.0325
GO:0032926	Negative regulation of activin receptor signaling pathway	2/10	0.0325
PRIMPOL$^{-/-}$			
GO:0000786	Nucleosome	12/108	4.56×10^{-14}
GO:0044815	DNA packaging complex	12/116	5.55×10^{-14}
GO:0032993	Protein-DNA complex	13/201	7.98×10^{-13}
GO:0006334	Nucleosome assembly	10/143	7.75×10^{-10}
GO:0031497	Chromatin assembly	10/162	2.17×10^{-9}
GO:0034728	Nucleosome organisation	10/174	3.69×10^{-9}
GO:0000785	Chromatin	14/525	4.41×10^{-9}
GO:0006333	Chromatic assembly or disassembly	10/187	5.67×10^{-9}
GO:0006323	DNA packaging	10/206	1.31×10^{-8}
GO:0065004	Protein-DNA complex assembly	10/246	6.75×10^{-8}

Table 24. An analysis of the GO terms significantly enriched in the genes deregulated between the knockout cell lines and the wildtype cells. GO terms with $p<0.05$ are shown unless there are greater than ten, in which case the top ten most significant hits are shown.

6.3.1.1 Comparisons of the knockout cell lines with wildtype cells cultured with G4 ligands in the undifferentiated state

To understand whether these genetic mutants were causing similar changes compared to culturing wildtype cells with G4 ligands in the undifferentiated state (from Results II), the deregulated genes were compared in the pluripotent state (Figure 107). This showed that there were very few overlaps between these conditions. The overlap was the greatest between PDS and PhenDC3 treated wildtype cells, suggesting that these changes were specific to ligand addition and not necessarily G4-specific. The overlapping genes with *REV1*$^{-/-}$ were *GAN*, *ENSG00000251095*, *TOMM40L*. With *WRN*$^{-/-}$ it was *MEF2C* and *PILRB* and with *PRIMPOL*$^{-/-}$ it was *GDF15* and *PILRB*. This data again supports the hypothesis that the undifferentiated cells are very capable of dealing with the G4 ligands as well as functioning without these enzymes.

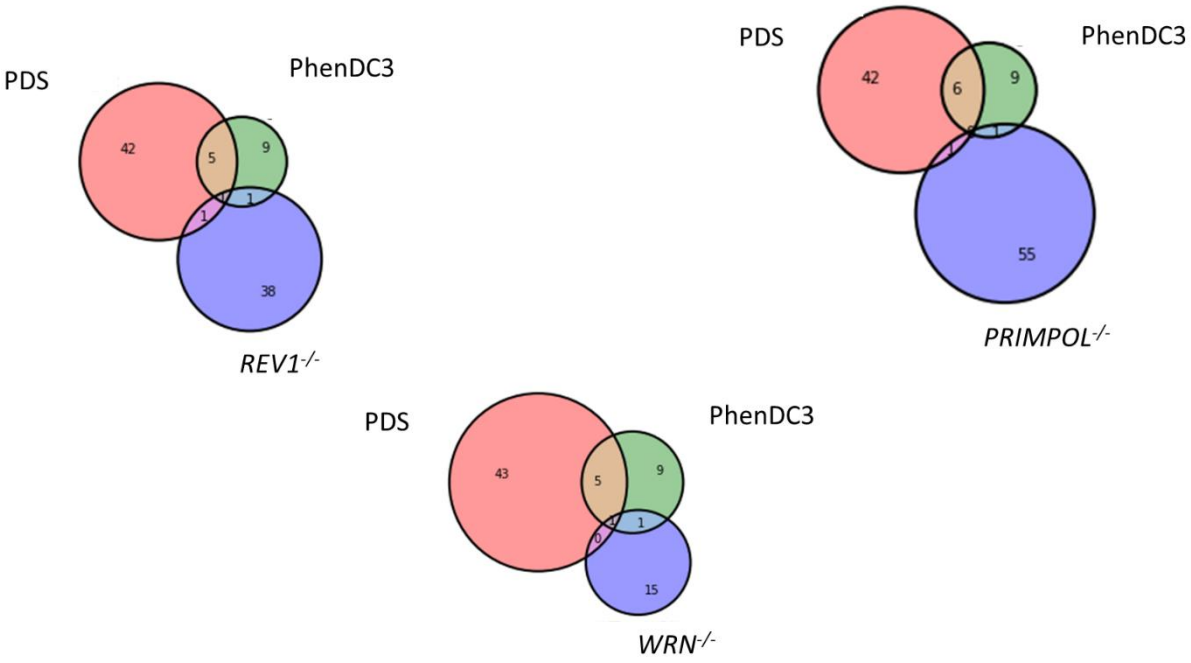


Figure 107. Venn diagrams to show overlapping genes deregulated in the undifferentiated state with the G4-binding ligands (Results III) compared with the knockout cell lines. *REV1*$^{-/-}$ is shown in the top left with only one gene deregulated in the three conditions, *WRN*$^{-/-}$ is shown on the bottom, again with only one shared gene and *PRIMPOL*$^{-/-}$ in the top right with no overlapping genes.

6.3.2 Gene expression changes during differentiation

6.3.2.1 Definitive endoderm differentiation of knockout cell lines

As was done in Section 6.3.1, RNA sequencing was performed in triplicate on the knockout cell lines *REV1*$^{-/-}$, *WRN*$^{-/-}$, *PRIMPOL*$^{-/-}$ and the wildtype cell line at 72 hours of definitive endoderm differentiation.

The deregulated gene expression was compared within each condition using Venn diagrams, as before (Figure 108) and the overlaps are shown in Table 3. As in the undifferentiated knockout cell line comparison, there were very few overlapping genes deregulated between each cell line. However, it was clear that the *REV1*$^{-/-}$ cells had a much higher number of genes deregulated in the differentiated state, suggesting a perturbed differentiation.

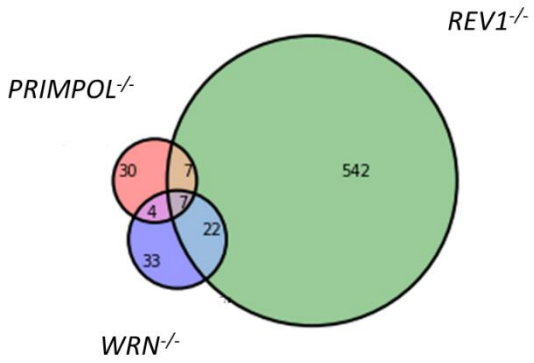


Figure 108. A Venn diagram to show the overlap of the deregulated genes in each of the knockout cell lines compared to the wildtype cell line. The names of the genes within each overlap are shown in Table 25. The *REV1*$^{-/-}$ cell line had the most perturbed gene expression but some genes were shared with *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$.

***REV1*$^{-/-}$ vs *PRIMPOL*$^{-/-}$**	***REV1*$^{-/-}$ vs *WRN*$^{-/-}$**	***PRIMPOL*$^{-/-}$ vs *WRN*$^{-/-}$**	***REV1*$^{-/-}$ vs *PRIMPOL*$^{-/-}$ vs *WRN*$^{-/-}$**
ADAMTSL5 SCN8A	*LAYN*	*ADAMTSL5 SCN8A*	*ADAMTSL5 SCN8A*
SLIT3	*FOXQ1*	*PIEZO1 RAB11FIP4*	*PID1*
CST1	*EDN1*	*PID1*	*URB1-AS1 LRRC61*
ZFP57	*CRHBP*	*URB1-AS1 LRRC61*	*GAN*
PID1	*SOX21*	*GAN*	*NOG*
URB1-AS1	*CD48*	*GRIK4*	
PYROXD2	*GAN*	*PLEKHA4*	
LRRC61	*SCN8A*	*NOG*	
GAN	*SNAI2*		
HIST1H2BJ	*PRUNE2*		
HIST1H2BK	*LRRC61*		
ATRX	*CCDC80*		
NOG	*COL7A1*		
	MARF1		
	CALB2		
	NOG		
	LEFTY2		
	IL11		
	PID1		
	NEAT1		
	ENSG00000237094		
	ADAMTSL5 SLC2A6		
	MAMDC2		
	TOMM40L		
	URB1-AS1 COL2A1		
	CDKN2B		
	NMI		

Table 25. A list of the overlap of deregulated genes in each knockout cell line at 72 hours of definitive endoderm differentiation, compared to the wildtype cell line. The largest comparison was seen in *REV1*$^{-/-}$ and *WRN*$^{-/-}$ with 29 genes.

The PCA of this data was generated (Figure 109) and this showed that the triplicates clustered well in each cell line at 72 hours. This PCA shows that the *REV1*$^{-/-}$ cell line showed the greatest expression difference compared to the wildtype. However, *PRIMPOL*$^{-/-}$ and *WRN*$^{-/-}$ show very little variance compared to the wildtype and therefore were able to differentiate to definitive endoderm with a similar efficiency.

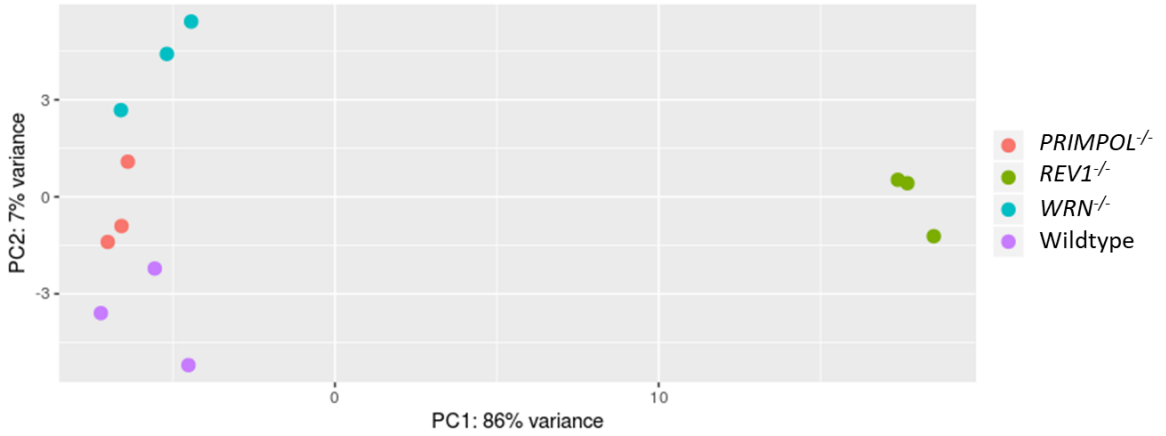


Figure 109. The PCA showing the variance of the wildtype and *REV1*$^{-/-}$, *WRN*$^{-/-}$ and *PRIMPOL*$^{-/-}$ cell lines at 72 hours of definitive endoderm differentiation. PC1 showed a much greater percentage of the variance, explaining why the *REV1*$^{-/-}$ cell line had the most different expression during differentiation. The *WRN*$^{-/-}$, *PRIMPOL*$^{-/-}$ and wildtype cell lines clustered individually but with very few changes between them.

6.3.2.2 Knockout cell line expression differences between undifferentiated and differentiated states

In order to compare the gene expression perturbations in each of the cell lines, the differences between the wildtype untreated and the knockout cell lines were compared in the undifferentiated and differentiated state: the deregulated gene lists were then contrasted within each mutant (Figure 110). The *REV1*$^{-/-}$ cell line had many more deregulated genes in the differentiated state, as seen in the PCA plot but there was a small overlap between these two states. Interestingly, *PRIMPOL*$^{-/-}$ cells did not have much change in the gene expression pattern in either state and there was a higher percentage of overlap of deregulated genes in both states. This suggested that this enzyme was required to similar levels in both the undifferentiated state and during differentiation. The overlapping genes are shown in Table 26. The *WRN*$^{-/-}$ cell line did not seem to alter gene expression significantly, possibly due to the known redundancy with *BLM* (Sarkies et al., 2012). The *PRIMPOL*$^{-/-}$ cell line had perturbed expression of a number of histone coding genes in both conditions, this was shared with the PhenDC3 G4 ligand treatment during differentiation, from Results III.

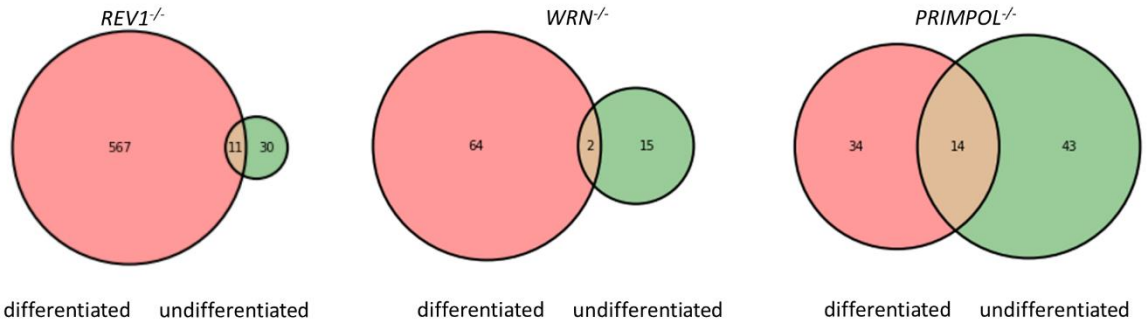


Figure 110. Venn diagrams to show the shared genes with deregulated expression in the undifferentiated state compared to 72 hours of definitive endoderm differentiation in each of the knockout cell lines. The highest similarity was seen in *PRIMPOL*$^{-/-}$ cells with 14 genes. A list of the genes is shown in Table 26.

REV1$^{-/-}$	***WRN***$^{-/-}$	***PRIMPOL***$^{-/-}$
RASL10A	*SCN8A*	*SCN8A*
ZFP57	*WRN*	*MED7*
TOMM40L		*LOC102724334*
URB1-AS1		*THG1L*
ENSG00000233056		*HMMR*
ENSG00000229618		*SLU7*
LRRC61		*HIST1H2BK*
GAN		*HIST1H2BJ*
FOXC1		*PTTG1*
ATRX		*MAT2B*
ENSG00000251095		*HIST1H1C*
		LINC02506
		ENSG00000255026
		IGLON5

Table 26. A list of deregulated genes shared by both the undifferentiated and differentiated state in each knockout cell line.

6.3.2.3 Comparison of gene expression profiles with PDS and PhenDC3 treated cells

The genes deregulated at 72 hours during the differentiation process can be compared between the PhenDC3 and PDS treated wildtype cells, and the *REV1*$^{-/-}$ cells (Figure 111). Interestingly there were 117 statistically significant differentially expressed genes shared by the PhenDC3, PDS and *REV1*$^{-/-}$ cells, and a greater number shared with just the PhenDC3 treated cells. The GO analysis is shown in Table 27. As these cells should all have G4-dependent gene expression changes, either due to the presence of ligands or the knockout of a processing enzyme, this suggests that the overlap of these genes is likely to be enriched in G-quadruplexes and this is likely to narrow down the search. On the other hand, there was no overlap between the three conditions with *PRIMPOL*$^{-/-}$ cells, and only four with *WRN*$^{-/-}$: *NEAT1*, *IL11*, *CD48* and *APOA1*.

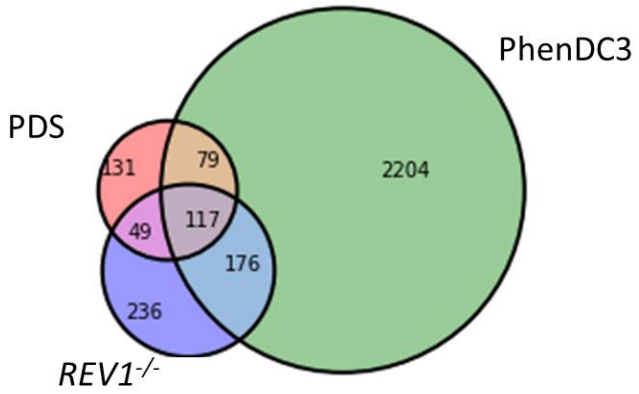


Figure 111. A Venn diagram to show the overlap of deregulated genes in ligand-treated wildtype cells compared to the *REV1*$^{-/-}$ cell line at 72 hours of definitive endoderm differentiation. 117 genes were common to all conditions.

GO ID	**Description**	**Count**	**Adjusted p-value**
GO:0032502	Developmental process	64/5686	4.25×10^{-8}
GO:0048856	Anatomical structure development	61/5342	1.77×10^{-7}
GO:0048514	Blood vessel morphogenesis	17/404	4.50×10^{-7}
GO:0030154	Cell differentiation	48/3642	7.49×10^{-7}
GO:0001568	Blood vessel development	18/486	8.42×10^{-7}
GO:0032501	Multicellular organismal process	69/6886	8.52×10^{-7}
GO:0048731	System development	53/4367	9.00×10^{-7}
GO:0001944	Vasculature development	18/509	1.74×10^{-6}
GO:0048869	Cellular developmental process	48/3734	1.81×10^{-6}
GO:0072358	Cardiovascular system development	18/519	2.36×10^{-6}

Table 27. The GO enrichment of the 117 genes common to PhenDC3 and PDS treatment, with the *REV1*$^{-/-}$ cell line at 72 hours of differentiation, shown in Figure 111.

6.3.3 Combining knockout cell lines with G4 ligands

While *REV1*$^{-/-}$ cells differentiated less efficiently than wildtype cells, differentiating in the presence of G4 ligands had not been tested, and this was likely to further perturb replication. *REV1*$^{-/-}$ cells were differentiated with 2 µM PDS and analysed using flow cytometry (Figure 112). Before analysis it was very clear that the cells were experiencing a greater level of cell death compared to the wildtype, and the data showed that the differentiation was incredibly inefficient, suggesting that this was very toxic to the cells: this may be by deregulating more genes containing G4s or by increasing DNA damage in the cells. This was repeated using 2 µM

PhenDC3, shown in Figure 113, and while there was a decrease in the efficiency of differentiation it was not as pronounced as with PDS. In both of these experiments, it was clear that *REV1*$^{-/-}$ cells were differentiating less efficiently than they had been previously, and this appeared to correlate with the length of time the cells had been cultured for. This is discussed in the following section.

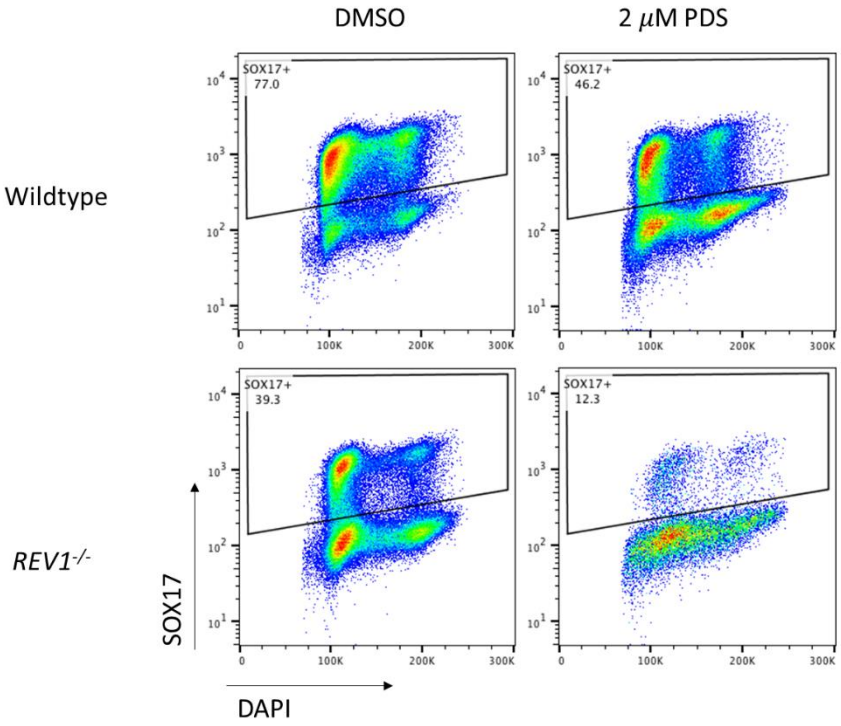


Figure 112. Flow cytometry data to show the differentiation at 72 hours in the wildtype and *REV1*$^{-/-}$ cell lines in the presence of 2 μM PDS. The efficiency is measured as the percentage of SOX17 positive cells at 72 hours, measured on the y-axis. The DNA content is measured on the x-axis.

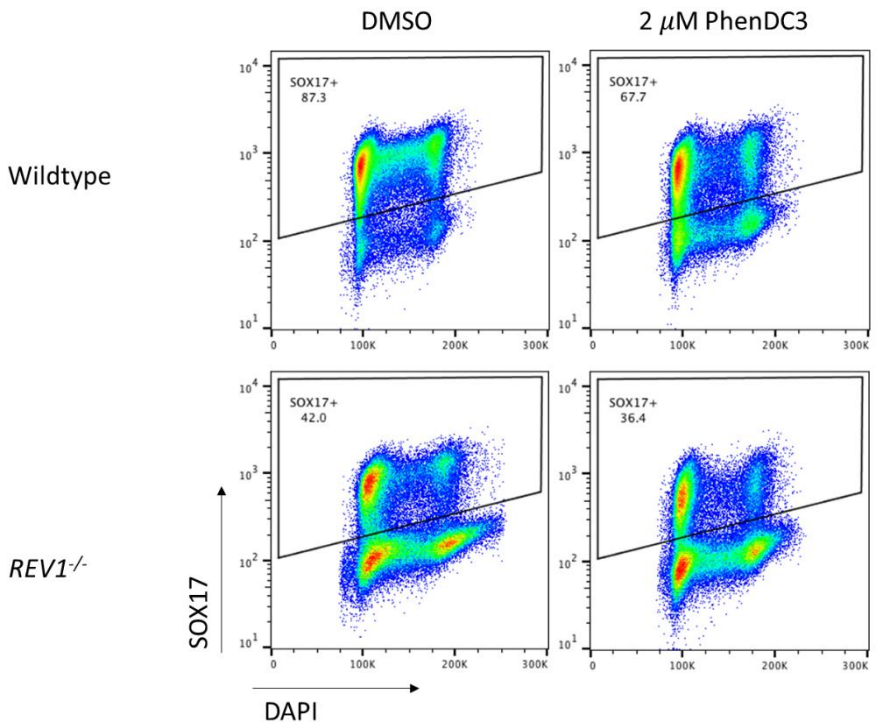


Figure 113. As in Figure 112 but with 2 µM PhenDC3.

6.4 Accumulation of changes over time

The knockout cell lines were differentiated to endoderm numerous times and it became clear that the longer the cells had been maintained in culture, the less efficient the differentiation became in the *REV1*$^{-/-}$ cell line specifically (Figure 114). The proportion of SOX17 positive cells decreased dramatically over the course of these experiments. Conveniently, the RNA sequencing was performed in cells cultured with a relatively high passage number and therefore the differentiation changes seen could be a build-up of epigenetic or genetic changes in the undifferentiated state. This was not the case in the undifferentiated state and may explain why fewer changes were seen. Ideally this would be repeated in cells cultured in the undifferentiated state for a longer period of time. This was incredibly interesting, and further work to analyse undifferentiated *REV1*$^{-/-}$ RNA over a number of passages would be very informative. It was also clear that the undifferentiated *REV1*$^{-/-}$ cells became sicker with each passage, and this may have contributed to the number of cells which had lost adherence, possibly perturbing the undifferentiated results discussed earlier. This is explained in further detail in the discussion.

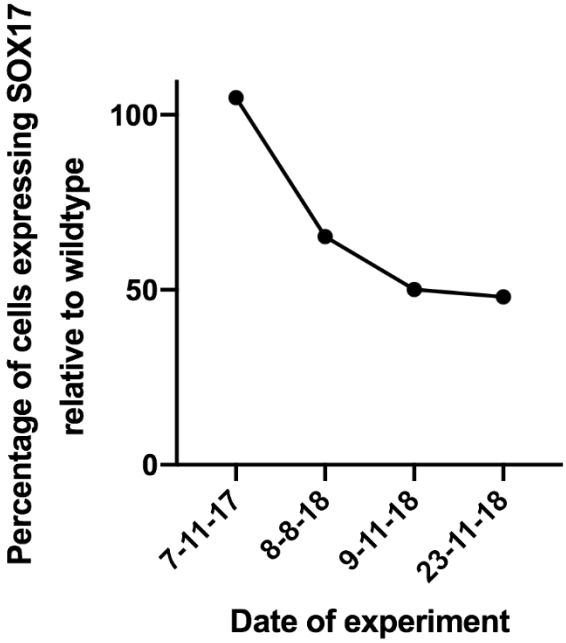


Figure 114. A graph to show the percentage of SOX17 positive *REV1*$^{-/-}$ cells compared to wildtype cells at 72 hours of definitive endoderm differentiation at four different points.

Chapter 7. Discussion

7.1 Initiation of the project

7.1.1 Previous work in the group

Previous work from the Sale group has focussed on the processing of secondary structures in DNA, exemplified by G-quadruplexes, and how the presence of such structures can alter DNA replication. The changes in DNA replication cause a local uncoupling of the replicative DNA helicase and polymerase which leads to long stretches of single stranded DNA, and the loss of marked nucleosomes associated with this region. This is able to cause a change in the epigenetic marks associated with this section of DNA and therefore cause gene expression changes (Sarkies et al., 2010, 2012). This phenomenon has been most intensively studied in the BU-1 locus of DT40 cells specifically using G4-binding ligands, genetic knockouts of enzymes involved in processing these structures, and nucleotide pool depletion (Guilbaud et al., 2017; Papadopoulou et al., 2015; Sarkies et al., 2012; Schiavone et al., 2014, 2016). However, there is no reason to limit the effect to G4 structures and a similar mechanism is proposed for R-loop structures in *PRIMPOL* deficient cells (Šviković et al., 2019). This model should hold for anything that uncouples the helicase and polymerase, causing a gap in processive replication, and is likely to include lesions in the DNA caused by damage.

All of the work used to generate the model was performed in DT40, a terminally differentiated chicken cell line, which may lack the ability to re-establish epigenetic states. Therefore, the initial aim of this thesis was to focus on similar mechanisms in a more biologically relevant system: differentiation. Since differentiating cells program their epigenome, it may be that they are robust to these perturbations. I used differentiation to understand whether impediments to DNA replication could alter the efficiency of differentiation with the larger aim of addressing whether development could be altered *in vivo*. The initial hypothesis was that efficiency of differentiation to definitive endoderm would be decreased on treatment with DNA damage or G4-binding ligands.

7.1.2 Human induced pluripotent stem cells

The BOBSC hiPS cell line was chosen due to the availability of the genetic knockout cell lines created in the Sanger Centre COMSIG project. The human iPSCs allowed the opportunity to study how ligands, DNA damaging agents and genetic knockout cell lines could alter the

undifferentiated state and the differentiation alone. The differentiation to definitive endoderm was chosen, with advice from the Vallier group. The protocol was very efficient, occurred over around five cell cycles and there was a good permeabilised flow cytometry read-out to monitor the efficiency on a single-cell basis, and a defined end point: the expression of SOX17. The gene expression changes in this protocol have been previously studied using microarrays (Teo et al., 2011). The differentiation to definitive endoderm was very reliable and was used throughout the thesis.

From using the G4 ligand PDC12 in this system, it was clear that DMSO alone was playing a role in decreasing the efficiency of differentiation without visibly altering the cell morphology. This had been previously noted during differentiation, and specifically during endoderm differentiation. One study showed that the differentiation to definitive endoderm could be improved by adding 0.5-0.6% DMSO in the early stages (Czysz et al., 2015), which is 50-fold higher than used for PhenDC3, PDS and NMM throughout this thesis. A different group showed that adding 2% DMSO in differentiation increased the differentiation to all lineages by activating Rb and increasing the proportion of cells in G1 phase (Chetty et al., 2013). Another study trialled a range of concentrations of DMSO during definitive endoderm differentiation and showed that medium and high doses of DMSO, the equivalent of 10 µM and 100 µM of the drugs used in this thesis, downregulated the level of SOX17 and high doses inhibited GATA6 expression altogether (Pal et al., 2012). Interestingly for this thesis, DMSO is used in PCR reactions to remove DNA secondary structures due to its high polarity and high dielectric constant. Therefore, it may be that DMSO is able to remove DNA secondary structures in the BOBSC cells, and this may be how it is acting in impacting differentiation. This could imply that DNA secondary structures are required during differentiation, but there is no further evidence to this conclusion.

The initial hypothesis was that differentiation in the presence of replication impediments would decrease the efficiency, monitored by the percentage of cells expressing SOX17 at 72 hours, by altering the expression of lineage specifying genes.

7.2 The response to DNA damage during definitive endoderm differentiation

7.2.1 Differences in the response to damage in the undifferentiated state compared to during differentiation

One of the first observations was that the undifferentiated and differentiating cells were able to tolerate differing quantities of DNA damaging agents. This was shown by increased cell death, a block in the cell cycle at G2/M and much greater gene expression changes as monitored using RNA sequencing, after treatment with DNA damaging agents during differentiation compared to the undifferentiated state. This could be due to differences in the amount of damage experienced, different responses to the same amount of damage, or a difference in the kinetics of repair. The cells could be experiencing different levels of DNA damage with the same concentration of damaging agent due to differences in the chromatin structure (Falk et al., 2008) which occur during differentiation. In order to understand whether the two cell states experience different numbers of breaks after DNA damage, an alkaline comet assay should be performed, and the olive tail moment compared in the undifferentiated state and during differentiation: this experiment is being performed as further work. Analysis of the different levels of damage could also be performed through monitoring CPDs and 6-4PPs induced by UV damage using ChIPseq.

There could also be a difference in response to the damage in each state which may be governed by the cell cycle, the expression of repair factors, and differences in the DNA damage response which are discussed throughout this thesis. Some of the differences seen between the undifferentiated and differentiating states are likely to be due to changes in the cell cycle during differentiation. ES cells were thought to have no G1 restriction point to allow the cell cycle to complete quickly, and to control the pluripotent state (Pauklin et al., 2011). Due to the lack of this restriction point, cells were thought to only respond to damage at the G2/M checkpoint (Neganova et al., 2011). The length of the cell cycle increases during differentiation and the cells gain this checkpoint, possibly causing the different responses to damage. However, since mESCs cultured in LIF and 2i have been shown to have an active G1 checkpoint (ter Huurne et al., 2017), this may not be the cause of the differences in response to damage between the states.

In this thesis, undifferentiated cells were shown to express high levels of DNA damage response proteins including p53, CHK1 and CHK2, and the levels decreased on differentiation to definitive endoderm. These proteins were not phosphorylated, suggesting they were not activated by the DNA damage response but were maintained at high levels. It may be that these

high levels allow the cells to remain poised for dealing with damage, suggesting why these pluripotent cells are better able to process damage.

Using the RNA sequencing data, it was seen that the level of transcripts of proteins implicated in DNA repair also changed during differentiation, these included *RPA2*, *PARP1* and *53BP1*. Interestingly, *53BP1* increases during differentiation which suggests that pathway choice occurs at this point between HR and NHEJ. This suggests an increased use of NHEJ during differentiation and may contribute to the differences in the cell response to damage.

7.2.2 Phosphorylation of H2A.X during unperturbed differentiation

In this thesis a major spike of phosphorylation of H2A.X at serine-139 was seen between 40 and 56 hours in unperturbed endoderm differentiation. This is likely to be at the point of the EMT during the differentiation to definitive endoderm. The EMT has been comprehensively studied, especially in the context of cancer. Previous work has described ATM phosphorylation of H2A.X which is required to induce HMGA2 transcription (Singh et al., 2015). This phosphorylation acts to impair H1 from binding, thus destabilising the nucleosome (Li et al., 2010). The phosphorylation of H2A.X at the EMT is associated with mass transcriptional activation caused by transcription factor binding (Singh et al., 2015). Some data has suggested that the binding of transcription factors is able to induce a sufficient replication fork blockage to induce DNA damage (Xia et al., 2019), however the spike seen in my data is shown mainly in G1 and G2, not S phase. In this thesis it is likely that the phosphorylation seen is associated with the EMT and large transcriptional changes associated with this transition. However, the *ATM*$^{-/-}$ cell line still showed phosphorylation at this time suggesting that if ATM does phosphorylate H2A.X then it is by a redundant mechanism, possibly compensated for by ATR, DNA-PK or another protein kinase. This mechanism is likely to be separate from the canonical DNA damage response. Ideally, I would follow individual cells to see if all cells which phosphorylated H2A.X committed to differentiate, but this would require live cell imaging of a phosphor-protein variant which is not currently possible.

Experiments performed in this thesis using inhibitors to attempt to address the question of which kinase was phosphorylating H2A.X were inconclusive. Most of the inhibitors induced damage and prevented efficient specification to definitive endoderm. While this could suggest that phosphorylation is required for differentiation, the knockout mouse model is inconsistent with this hypothesis; mice with a genetic knockout of *H2AFX* are viable but show increased genetic instability (Celeste et al., 2002). Further work would include treatment with caffeine, a non-

specific inhibitor of ATM and ATR, to see whether this prevented the phosphorylation. Conditional knockout of *ATR* and *PRKDC* in combination with *ATM* could help to address this question but since a single knockout *ATR* cell line is not viable, this is unlikely to be a practical solution. It would also be interesting to follow the dephosphorylation of H2A.X after damage as this has been proposed to occur on the nucleosome and not occur as a result of histone turnover (Chowdhury et al., 2005; Nazarov et al., 2003; Siino et al., 2002).

It will be important to understand the role of the phosphorylation of H2A.X at the EMT and whether it is related with the DNA damage response in any way. Analysis of the ChIP data should allow a greater understanding of the regions of DNA these marks are associated with, specifically promoters of EMT expressed genes. In order to rule out that this mark is associated with DNA damage it is also important to analyse more DDR markers at this point in differentiation. Using immunofluorescence would allow analysis of co-localisation of these markers, such as 53BP1 and γH2A.X. It is also important to verify this antibody, further work analysing the levels of total H2A.X will be required to rule out the possibility that it is the total protein changing.

7.2.3 The result of inflicting DNA damage during differentiation

DNA damaging agents were added during differentiation to induce damage and analyse the course of differentiation after treatment. MMS, HU and UV treatment prevented efficient differentiation, as measured by SOX17 expression, in a dose dependent manner. All treatment induced an increase in cells in G2/M phase of the cell cycle, indicative of DNA damage, likely to be due to checkpoint activation. A cell cycle block and increased cell death occurred to a much higher level during differentiation than in the undifferentiated state with the same treatment.

Treating with MMS did decrease the number of cell cycles gone through during differentiation slightly, probably due to the G2/M block slowing the cell cycle. However, this difference was not enough to cause the decrease in SOX17 expression because SOX17 protein could be seen from around three cell cycles and tended to be present to an extent in wildtype cells from around 40 hours. This suggested that there was not a trivial reason for this difference in expression and this was highlighted by showing that increasing the time of differentiation did not increase the proportion of SOX17 expressing cells.

An effect was seen when damage was induced at any point during differentiation, but the greatest decrease in SOX17 was at around 24 hours. The lower response after 48 hours may be due to the cells already becoming specified to endoderm and having turned on SOX17 expression.

This window of time after specification of differentiation but before expression of SOX17 may suggest that more damage was experienced by the cells at around 24 hours, possibly also due to changes in chromatin structure during differentiation. This time may correlate with the point in which the levels of DDR proteins decrease during differentiation. The resulting upregulation of DDR proteins after treatment with DNA damaging agents may prevent further differentiation in these cells. This window of time may also correlate with cell fate choice and lineage commitment pathway decisions. The outcome could be affected by the position of the cells in the cell cycle when damage was induced, as the cell cycle distribution was altered throughout differentiation.

Cell synchronisation would allow DNA damaging agents to be added at set points in the cell cycle and it would be interesting to see whether the phase that a cell was in would alter the differentiation efficiency.

7.2.4 The role of p53 in controlling differentiation

Previous work has shown that p53 is redundantly required for induction of mesendoderm differentiation (Wang et al., 2017), explaining why it is kept at a high level at the onset of endoderm differentiation. p53 has also been shown to play a role in the EMT by initiating the transition (Chang et al., 2011). In the epithelial state p53 protein levels are high and this is required for controlling expression of EMT specific genes: overexpression of p53 in the mesenchymal state reverts the phenotype to epithelial. Therefore, p53 is needed to initiate the EMT program and is then downregulated by MDM2 to allow completion of this transition (Araki et al., 2010). This potentially explains the observations in this thesis that p53 is downregulated during differentiation, specifically at the time that H2A.X is phosphorylated, further highlighting that this is associated with the EMT.

Differentiation in the *TP53*$^{-/-}$ cell line occurred in a similarly efficient manner to the wildtype cells. This was expected because genetic knockout of *Trp53* in mice results in normal differentiation most likely because the other family members are able to compensate (Donehower et al., 1992). The *TP53*$^{-/-}$ cell line was used to understand the role played by p53

in the presence of DNA damaging agents during differentiation. This cell line showed less cell death during unperturbed differentiation than the wildtype cell line, possibly suggesting less surveillance of damage in these cells. This could lead to increased problems in embryogenesis but as this *in vitro* system is very different from the *in vivo* situation this may not have any relevance. When DNA damage was induced during differentiation, the *TP53*$^{-/-}$ cell line differentiated efficiently, similarly to untreated cells but unlike wildtype cells. Therefore, it is likely that *TP53*$^{-/-}$ treated cells accumulate damage that may manifest later in the developed organism.

However, there was still a large G2/M block in the cell cycle in treated *TP53*$^{-/-}$ cells. This was interesting because it was clearly not the accumulation of cells in G2/M phase that was preventing differentiation. The cell cycle phase has been shown to alter the propensity to differentiate (Gonzales et al., 2015; Pauklin and Vallier, 2013; Pauklin et al., 2016) but these studies were looking at the initiation of differentiation and the escape from pluripotency. Therefore, this may not be the case when the DNA damage is induced after commitment has occurred. The cell cycle block is likely to be caused by ATM phosphorylating p53, H2A.X and CHK2 and inhibiting cdc25, thus preventing activation of CDK1 and movement through the cell cycle. This can also occur through the ATR CHK1 pathway.

7.2.5 RNA sequencing analysis of treatment with DNA damaging agents in the undifferentiated state and during differentiation

The RNA sequencing was designed such that each of the triplicates were collected on the same day. The caveats of this design are that any confounding variables are not accounted for. However, since the efficiency of differentiation was checked by flow cytometry prior to performing library preparation, this effect should be minimised. The data was analysed using the Cufflinks RNAseq pipeline for comparison to other work performed in the group but prior to publishing it will be interesting to also analyse this data using the DESeq2 pipeline. In order to rule out any variables the triplicates should ideally be collected on different days to factor in different batches of media, different times of the incubator being open and difference in seeding density. There are a number of biases to be aware of when analysing sequencing data because they can lead to very significant p-values in data enrichment analyses without any biological significance. Technical biases can arise due to base bias derived from different polymerases used in the PCR reaction. This bias occurs as some polymerases preferentially catalyse elongation over A/T-rich or G/C-rich regions, this is important when using the G4 ligands in the next section. A high GC bias can manifest itself as the GO term ‘DNA-Templated

Transcription' $p=2x10^{-20}$ in terms of the significant genes, whereas a low GC bias can lead to 'GPCR Signalling' $p=4x10^{-12}$ (unpublished, Babraham Bioinformatics Department, Cambridge). There are also statistical biases including the length of a gene; it is harder to see changes in shorter genes. This can again, lead to a highly significant GO analysis: 'synapse' $p=2x10^{-30}$ (*Ibid*). There are also some gene sets that have common hits for non-specific reasons: ribosomal, cytoskeleton, extracellular, secreted, translation. These biases are important to consider when analysing data of this nature.

Treatment with MMS in the undifferentiated state induced more changes in wildtype cells compared to *TP53*$^{-/-}$, as expected, but there were very few changes in either. This was in stark contrast to treatment during differentiation. Comparing the wildtype and *TP53*$^{-/-}$ cells treated with damage, it was clear that the deregulated pathways were involved in apoptosis, DNA damage, and checkpoint activation, as expected, further suggesting that without p53 the cells cannot respond to damage in the same manner.

During differentiation, it was clear that adding MMS to wildtype cells altered differentiation as the most significant GO term enrichment was pathways involved in morphogenesis and development. However, this was not the case in *TP53*$^{-/-}$ cells, again suggesting that inducing damage did not cause changes to differentiation in these cells. This was further seen by the number of deregulated *TP53*$^{-/-}$ transcripts decreasing from 48 to 72 hours, after the damage. In the wildtype cells the magnitude of deregulated genes was much greater and increased during differentiation, suggesting that the cells were differentiating down a different pathway. This enhanced the view of the flow cytometry experiments, but further analysis of this sequencing data will allow a much more detailed understanding of this process. Further analysis of the expression of genes expressed in each lineage will allow an understanding of whether treatment simply prevents differentiation or directs differentiation towards a different lineage such as mesoderm.

7.2.6 Further work to understand the mechanism of the DDR in differentiation

A lot of information has been gained from the BOBSC knockout Sanger Centre COMSIG cells, including the *REV1*$^{-/-}$ and *TP53*$^{-/-}$ cells. Using *CHEK2*$^{-/-}$ and *H2AFX*$^{-/-}$ would be informative as to the method of differentiation inhibition, the EMT and cell cycle inhibition. The *MDM2*$^{-/-}$ cells would also allow understanding of whether preventing downregulation of p53 is able to inhibit differentiation.

Although the BOBSC cell line has been helpful in this project, it contains a translocation which may alter the generality of these findings. Therefore, it will be important to confirm these findings in the untranslocated BOBSC cell line, and other iPSCs. Repeating the work in the H9 and other hESC lines will be very helpful. However, there are likely to be some changes due to the differences in these cell lines and their different genetic backgrounds.

7.3 The ability of the G4 secondary structure to alter differentiation

7.3.1 G4-binding ligands alter the fate of differentiation

Both PDS and PhenDC3 treatment were decreasing the expression of SOX17 significantly at 48 and 72 hours but not the initial expression of EOMES at 24 hours. The changes to SOX17 showed that definitive endoderm differentiation was perturbed. As with DNA damage, the ligands needed to be present during differentiation and not added after, in which case no decrease in SOX17 expression was observed. This further suggests that endoderm specification must be perturbed before commitment, but after 72 hours this can no longer be altered. However, the action of PDS could be prevented by using the genetic knockout *TP53*$^{-/-}$ cell line and therefore the effect of DNA damage outweighed any G4-specific replication changes. The reason behind PhenDC3 not triggering DNA damage or a DDR in the cell is not yet known but is likely to be due to the method of binding to the G4 and the kinetics of release.

Altering the point of PhenDC3 treatment suggested that addition for the first 24 hours caused the greatest effect on the expression of SOX17. This implied that PhenDC3 was likely to be causing gene expression changes early on which then propagate to greater changes at the end of differentiation. Preliminary media switching experiments in PhenDC3 treated cells showed that secreted factors produced in the mock-treated medium in the first six hours could at least partially compensate for the addition of PhenDC3.

If this outcome was due to a G4 being stabilised by PhenDC3 and inducing a replication impediment, then this should occur stochastically with each cell cycle. Therefore, it is assumed that the longer the cells are cultured in the presence of ligand, the greater the change. If the ligand were interfering with a transcription factor binding site then this would cause dose-dependent changes to the equilibrium. However, during development there are key windows where small changes in gene expression will affect the outcome overall. This data does not show whether this is a replication dependent epigenetic change or whether the ligand interferes

with transcription factor binding patterns, or other methods of regulation, but it is likely to happen early in differentiation.

Checking lineage specific markers of mesoderm, it was clear that both PDS, especially at early timepoints, and PhenDC3 treated cells were expressing markers of mesoderm differentiation. This suggested that the drugs were not preventing initial differentiation to mesendodermal fate, but the cells were not committing to an endodermal fate. Surprisingly, this increase in mesoderm specification genes was shared with MMS treated wildtype cells. This fate choice suggests that when differentiation is perturbed the cells choose mesodermal fate in preference to endodermal, in this system.

Due to the fact that both G4 ligands were not inducing identical changes, it is hard to suggest a mechanism to cause these changes specific to this secondary structure. The PDS was inducing DNA damage and acted in part similarly to MMS. However, PhenDC3 was not inducing DNA damage and the expression profile was different to the PDS treatment.

7.3.2 *REV1* knockout cells show altered definitive endoderm differentiation

REV1$^{-/-}$ cells showed very different gene expression compared to wildtype cells using RNAseq. This was interesting because in the initial experiments the differentiation efficiency had been as high as in the wildtype cells. However, observations showed that there were increased numbers of cells losing adherence, and less rounded colonies in the *REV1*$^{-/-}$ cells in the undifferentiated state and this was exacerbated the longer the cells were kept in culture. The decrease in the ability to differentiate also seemed to occur with an increase in passage number. This could be of significance because it may be consistent with long term culture leading to epigenetic alterations and changes in the transcriptome over time, similar to the observations made in *REV1*$^{-/-}$ DT40 cells.

7.3.3 Similarities between the *REV1* knockout cell line and G4 ligand treatment during definitive endoderm differentiation

There was a high level of overlap in the genes deregulated on ligand treatment and the *REV1*$^{-/-}$ cell line at 72 hours of endoderm differentiation. The enrichment analysis for this list showed GO terms associated with differentiation and developmental processes with significant p-values. Further work suggested that combining the *REV1*$^{-/-}$ cell line with the G4 ligands further decreased the efficiency of differentiation, suggesting that this could exacerbate the phenotype further. The mechanism may be that the drugs are increasing the number of G4s further than

biologically relevant, and *REV1*$^{-/-}$ cells struggle further to process them. This suggests that G4s may be able to alter differentiation specifically, and further analysis of the enrichment of predicted G4 structures is being performed.

7.3.4 Further sequencing data analysis

While a general analysis of all of the sequencing data in more detail will provide greater insight into the pathway perturbations and the differences between the ligands and the DNA damaging agents, there will also be more specific analysis required to understand whether these PhenDC3 and *REV1*$^{-/-}$ induced changes are G4-specific. A greater metanalysis of G4 or G-rich regions and their proximity to introns, TSSs, transcription factor binding sites and other features would be informative to show if there is an association, as previously performed by other groups (Eddy et al., 2011; Tang et al., 2016). It would also be interesting to attempt to find a general consensus for the different types of G4 bound by each drug, if either had a preference for binding to a given structure.

The initial idea proposed was to try to narrow down a subset of genes that PhenDC3 acts on, but having analysed the data, this looks to be an unrealistic goal. However, more recombinant protein experiments, such as GSC addition, may help to find some factors. Further RNAseq on very early timepoints on PhenDC3 treatment may help to elucidate the earlier expression changes that may alter further pathways. Single-cell RNAseq may also allow an understanding of whether these processes are stochastic, and therefore whether they are likely to act in a similar manner to BU1 loss in DT40. The final aim still remains to narrow down a list of G4s that may be mechanistically responsible for these changes.

The thesis is missing any knowledge of the underlying chromatin, and changes to this with the treatments and during differentiation. Therefore, it would be interesting to perform ATACseq and ChIPseq using the ligands, knockouts and damaged cells to understand how treatment are able to alter the epigenetic state at loci associated with changes during this differentiation.

7.4 Conclusions

I set out to address the question ‘is differentiation robust to replication impediments’ but this work has uncovered two separate issues: the effect of the DDR on differentiation and the possible direct effect of G4s on differentiation.

In conclusion, this thesis has shown that there are vast changes between cells cultured in the pluripotent state and those undergoing differentiation. This occurs at the level of the transcriptome, cell cycle and response to DNA damage. While much work has been done on comparing undifferentiated cells and differentiated cells, far less has been performed on the comparison to differentiating cells, in these early lineages.

I propose that G4 ligands are able to alter the course of definitive endoderm differentiation via two separate means. PDS prevents endoderm specification through the DDR whereas PhenDC3 induces gene expression changes, altering a vast array of pathways, and this stops movement through the EMT without a significant activation of the DDR. Further work will allow a more in-depth analysis of the possible G4s responsible for these changes. Interestingly, change in the expression of SOX17 during differentiation is stochastic in PhenDC3 treated cells, potentially supporting a replication-dependent phenomenon.

I suggest that during differentiation, levels of DDR proteins decrease, controlled either at the level of the RNA or protein depending on the factor. This alters the response to DNA damaging agents, as well as G4 ligands, and can induce increased levels of cell death. During differentiation, H2A.X is phosphorylated at the EMT, most likely by ATM but this may be redundant, and all other proteins phosphorylated by ATM have low expression at this time so that they are not phosphorylated by ATM and do not induce the canonical DDR. However, when damage is induced, these proteins are upregulated and become phosphorylated by ATM, inducing the DDR and preventing differentiation, I have termed this a ‘differentiation checkpoint’ (Figure 115). When p53 is not present this checkpoint does not occur and differentiation continues in the face of damage, possibly leading to increased damage in the cell. This checkpoint may be a direct transcriptional effect of p53 turning on mesoderm specifying genes or acting via the canonical DDR but this remains to be elucidated.

Despite vast amounts of work studying the way a cell responds to damage in differentiated cells, the response during differentiation has barely been touched on. Changes in the DNA damage response during differentiation are very important to understand the reaction to endogenous and exogenous damage during differentiation. This fundamental understanding is interesting and has rarely been studied.

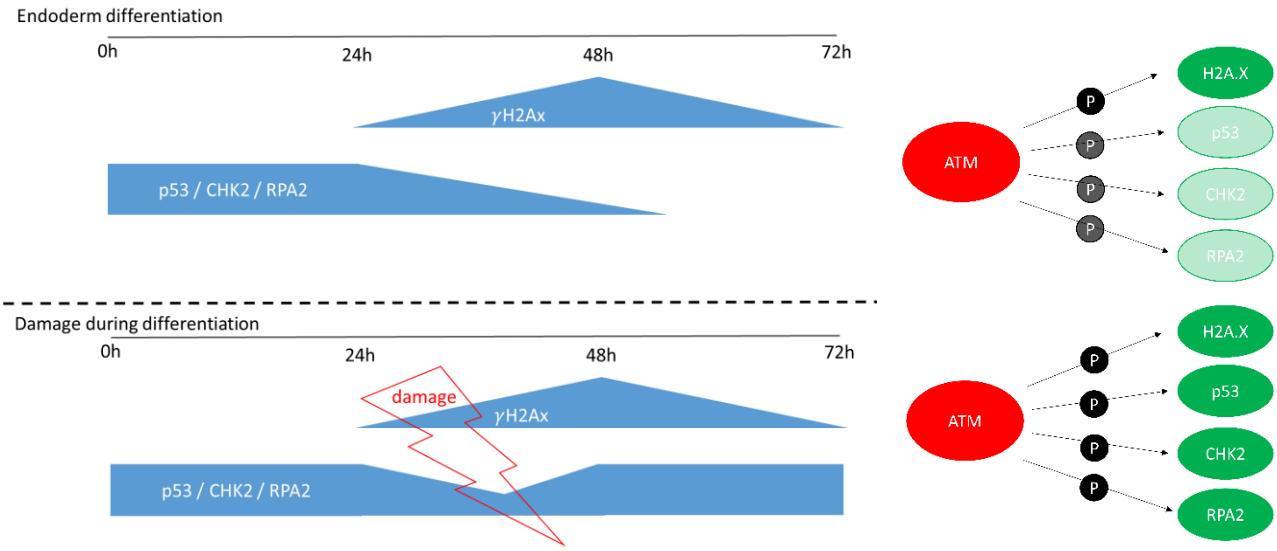


Figure 115. A model to suggest how DNA damage induces a 'differentiation checkpoint'. The expression of DDR factors decreases during differentiation and H2A.X is phosphorylated at 48 hours by ATM. However, when damage is induced, p53 and possibly other DDR proteins are upregulated and ATM is also able to phosphorylate these proteins. The phosphorylation of p53 causes this differentiation checkpoint to become active and prevent endoderm specification.

Bibliography

Adamo, A., Sesé, B., Boue, S., Castaño, J., Paramonov, I., Barrero, M.J., and Belmonte, J.C.I. (2011). LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. Nat. Cell Biol. *13*, 652–659.

Agaronyan, K., Morozov, Y.I., Anikin, M., and Temiakov, D. (2015). Replication-transcription switch in human mitochondria. Science (80-.). *347*, 548–551.

Ahuja, A.K., Jodkowska, K., Teloni, F., Bizard, A.H., Zellweger, R., Herrador, R., Ortega, S., Hickson, I.D., Altmeyer, M., Mendez, J., et al. (2016). A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. Nat. Commun. *7*, ncomms10660.

Alvino, G.M., Collingwood, D., Murphy, J.M., Delrow, J., Brewer, B.J., and Raghuraman, M.K. (2007). Replication in Hydroxyurea: It's a Matter of Time. Mol. Cell. Biol. *27*, 6396–6406.

Araki, S., Eitel, J.A., Batuello, C.N., Bijangi-Vishehsaraei, K., Xie, X.-J., Danielpour, D., Pollok, K.E., Boothman, D.A., and Mayo, L.D. (2010). TGF-β1–induced expression of human Mdm2 correlates with late-stage metastatic breast cancer. J. Clin. Invest. *120*, 290–302.

Arama, E., Agapite, J., and Steller, H. (2003). Caspase activity and a specific cytochrome C are required for sperm differentiation in Drosophila. Dev. Cell *4*, 687–697.

Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc. Natl. Acad. Sci. *88*, 10148–10152.

Arora, A., and Suess, B. (2011). An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. RNA Biol. *8*, 802–805.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Baba, T.W., Giroir, B.P., and Humphries, E.H. (1985). Cell lines derived from avian lymphomas exhibit two distinct phenotypes. Virology *144*, 139–151.

Banáth, J.P., Bañuelos, C.A., Klokov, D., MacPhail, S.M., Lansdorp, P.M., and Olive, P.L. (2009). Explanation for excessive DNA single-strand breaks and endogenous repair foci in pluripotent mouse embryonic stem cells. Exp. Cell Res. *315*, 1505–1520.

Banin, S., Moyal, L., Shieh, S., Taya, Y., Anderson, C.W., Chessa, L., Smorodinsky, N.I., Prives, C., Reiss, Y., Shiloh, Y., et al. (1998). Enhanced phosphorylation of p53 by ATM in response to DNA damage. Science *281*, 1674–1677.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Barlow, C., Hirotsune, S., Paylor, R., Liyanage, M., Eckhaus, M., Collins, F., Shiloh, Y., Crawley, J.N., Ried, T., Tagle, D., et al. (1996). Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell *86*, 159–171.

Beaudoin, J.-D., Jodoin, R., and Perreault, J.-P. (2014). New scoring system to identify RNA G-quadruplex folding. Nucleic Acids Res. *42*, 1209–1223.

Bedrat, A., Lacroix, L., and Mergny, J.-L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. *44*, 1746–1759.

Beranek, D.T. (1990). Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. Mutat. Res. *231*, 11–30.

van den Berg, D.L.C., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2008). Estrogen-Related Receptor Beta Interacts with Oct4 To Positively Regulate Nanog Gene Expression. Mol. Cell. Biol. *28*, 5986–5995.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., and Lemaitre, J.-M. (2012). Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat. Struct. Mol. Biol. *19*, 837–844.

Bétous, R., Rey, L., Wang, G., Pillaire, M.-J., Puget, N., Selves, J., Biard, D.S.F., Shin-ya, K., Vasquez, K.M., Cazaux, C., et al. (2009). Role of TLS DNA polymerases eta and kappa in processing naturally occurring structured DNA in human cells. Mol. Carcinog. *48*, 369–378.

Biebricher, A., Hirano, S., Enzlin, J.H., Wiechens, N., Streicher, W.W., Huttner, D., Wang, L.H.-C., Nigg, E.A., Owen-Hughes, T., Liu, Y., et al. (2013). PICH: A DNA Translocase Specially Adapted for Processing Anaphase Bridge DNA. Mol. Cell *51*, 691–701.

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. *5*, 182–186.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.
Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. *13*, 770–780.

Boeuf, H., Hauss, C., Graeve, F. De, Baran, N., and Kedinger, C. (1997). Leukemia Inhibitory Factor–dependent Transcriptional Activation in Embryonic Stem Cells. J. Cell Biol. *138*, 1207–1217.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-

cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. *16*, 513–525.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell *122*, 947–956.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Brown, E.J., and Baltimore, D. (2000). ATR disruption leads to chromosomal fragmentation and early embryonic lethality. Genes Dev. *14*, 397–402.

Brown, E.J., and Baltimore, D. (2003). Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. Genes Dev. *17*, 615–628.

Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q.-L., and Smith, A. (2008). Capture of Authentic Embryonic Stem Cells from Rat Blastocysts. Cell *135*, 1287–1298.

Buerstedde, J.M., and Takeda, S. (1991). Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell *67*, 179–188.

Bugaut, A., and Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. *40*, 4727–4741.

Burley, G.A., Dominguez, C., Weldon, C., Hurley, L.H., Eperon, I.C., Branlant, C., Boddupally, P.V.L., Dacanay, J.G., Gokhale, V., and Behm-Ansmant, I. (2017). Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. Nucleic Acids Res. *46*, 886–896.

Burrow, A.A., Marullo, A., Holder, L.R., and Wang, Y.-H. (2010). Secondary structure formation and DNA instability at fragile site FRA16B. Nucleic Acids Res. *38*, 2865–2877.

Byun, T.S., Pacek, M., Yee, M., Walter, J.C., and Cimprich, K.A. (2005). Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. Genes Dev. *19*, 1040–1052.

Calder, A., Roth-Albin, I., Bhatia, S., Pilquil, C., Lee, J.H., Bhatia, M., Levadoux-Martin, M., McNicol, J., Russell, J., Collins, T., et al. (2013). Lengthened G1 Phase Indicates Differentiation Status in Human Embryonic Stem Cells. Stem Cells Dev. *22*, 279–295.

Campbell, K.H.S., McWhir, J., Ritchie, W.A., and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. Nature *380*, 64–66.

Canman, C.E., Lim, D.S., Cimprich, K.A., Taya, Y., Tamai, K., Sakaguchi, K., Appella, E., Kastan, M.B., and Siliciano, J.D. (1998). Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. Science *281*, 1677–1679.

Capra, J.A., Paeschke, K., Singh, M., and Zakian, V.A. (2010). G-Quadruplex DNA

Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae. PLoS Comput. Biol. *6*, e1000861.

Catez, F., Yang, H., Tracey, K.J., Reeves, R., Misteli, T., and Bustin, M. (2004). Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. Mol. Cell. Biol. *24*, 4321–4328.

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Méchali, M. (2012). New insights into replication origin characteristics in metazoans. Cell Cycle *11*, 658–667.

Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J., et al. (2002). Genomic Instability in Mice Lacking Histone H2AX. Science (80-.). *296*, 922–927.

Cervantes, R.B., Stringer, J.R., Shao, C., Tischfield, J.A., and Stambrook, P.J. (2002). Embryonic stem cells and somatic cells differ in mutation frequency and type. Proc. Natl. Acad. Sci. U. S. A. *99*, 3586–3590.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chang, C.-J., Chao, C.-H., Xia, W., Yang, J.-Y., Xiong, Y., Li, C.-W., Yu, W.-H., Rehman, S.K., Hsu, J.L., Lee, H.-H., et al. (2011). p53 regulates epithelial–mesenchymal transition and stem cell properties through modulating miRNAs. Nat. Cell Biol. *13*, 317.

Chetty, S., Pagliuca, F.W., Honore, C., Kweudjeu, A., Rezania, A., and Melton, D.A. (2013). A simple tool to improve pluripotent stem cell differentiation. Nat. Methods *10*, 553–556.

Cheung, I., Schertzer, M., Rose, A., and Lansdorp, P.M. (2002). Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. Nat. Genet. *31*, 405–409.

Chia, C.Y., Madrigal, P., Denil, S.L.I.J., Martinez, I., Garcia-Bernardo, J., El-Khairi, R., Chhatriwala, M., Shepherd, M.H., Hattersley, A.T., Dunn, N.R., et al. (2019). GATA6 Cooperates with EOMES/SMAD2/3 to Deploy the Gene Regulatory Network Governing Human Definitive Endoderm and Pancreas Formation. Stem Cell Reports *12*, 57–70.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.
Chowdhury, D., Keogh, M.-C., Ishii, H., Peterson, C.L., Buratowski, S., and Lieberman, J. (2005). gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. Mol. Cell *20*, 801–809.

De Cian, A., Delemos, E., Mergny, J.-L., Teulade-Fichou, M.-P., and Monchaud, D. (2007). Highly efficient G-quadruplex recognition by bisquinolinium compounds. J. Am. Chem. Soc. *129*, 1856–1857.

Ciccia, A., and Elledge, S.J. (2010). The DNA Damage Response: Making It Safe to Play with Knives. Mol. Cell *40*, 179–204.

Cockayne, E.A. (1946). Dwarfism with retinal atrophy and deafness. Arch. Dis. Child. *21*, 52–54.

Cogoi, S., Shchekotikhin, A.E., and Xodo, L.E. (2014). HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. Nucleic Acids Res. *42*, 8379–8388.

Copp, A.J. (1979). Interaction between inner cell mass and trophectoderm of the mouse blastocyst. II. The fate of the polar trophectoderm. J. Embryol. Exp. Morphol. *51*, 109–120.

Cortez, D., Guntuku, S., Qin, J., and Elledge, S.J. (2001). ATR and ATRIP: Partners in Checkpoint Signaling. Science (80-.). *294*, 1713–1716.

Coucouvanis, E., and Martin, G.R. (1999). BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. Development *126*, 535–546.

Crabbe, L., Verdun, R.E., Haggblom, C.I., and Karlseder, J. (2004). Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. Science *306*, 1951–1953.

Crenshaw, E., Leung, B.P., Kwok, C.K., Sharoni, M., Olson, K., Sebastian, N.P., Ansaloni, S., Schweitzer-Stenner, R., Akins, M.R., Bevilacqua, P.C., et al. (2015). Amyloid Precursor Protein Translation Is Regulated by a 3'UTR Guanine Quadruplex. PLoS One *10*, e0143160.

Czechanski, A., Byers, C., Greenstein, I., Schrode, N., Donahue, L.R., Hadjantonakis, A.-K., and Reinholdt, L.G. (2014). Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. Nat. Protoc. *9*, 559–574.

Czysz, K., Minger, S., and Thomas, N. (2015). DMSO efficiently down regulates pluripotency genes in human embryonic stem cells during definitive endoderm derivation and increases the proficiency of hepatic differentiation. PLoS One *10*, e0117689.

D'Antonio, L., and Bagga, P. Computational methods for predicting intramolecular g-quadruplexes in nucleotide sequences. In Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004., (IEEE), pp. 561–562.

David, S.S., O'Shea, V.L., and Kundu, S. (2007). Base-excision repair of oxidative DNA damage. Nature *447*, 941–950.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

DeStephanis, D., McLeod, M., and Yan, S. (2015). REV1 is important for the ATR-Chk1 DNA damage response pathway in Xenopus egg extracts. Biochem. Biophys. Res. Commun. *460*, 609–615.

Ding, H., Schertzer, M., Wu, X., Gertsenstein, M., Selig, S., Kammori, M., Pourvali, R., Poon, S., Vulto, I., Chavez, E., et al. (2004). Regulation of Murine Telomere Length by Rtel. Cell *117*, 873–886.

Domaschenz, R., Kurscheid, S., Nekrasov, M., Han, S., and Tremethick, D.J. (2017). The

Histone Variant H2A.Z Is a Master Regulator of the Epithelial-Mesenchymal Transition. Cell Rep. *21*, 943–952.

Donehower, L.A. (1996). The p53-deficient mouse: a model for basic and applied cancer studies. Semin. Cancer Biol. *7*, 269–278.

Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Butel, J.S., and Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature *356*, 215–221.

Dong, D.W., Pereira, F., Barrett, S.P., Kolesar, J.E., Cao, K., Damas, J., Yatsunyk, L.A., Johnson, F., and Kaufman, B.A. (2014). Association of G-quadruplex forming sequences with human mtDNA deletion breakpoints. BMC Genomics *15*, 677.

Dovey, O.M., Foster, C.T., and Cowley, S.M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. Proc. Natl. Acad. Sci. *107*, 8242–8247.

Eddy, J., and Maizels, N. (2006). Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. *34*, 3887–3896.

Edmunds, C.E., Simpson, L.J., and Sale, J.E. (2008). PCNA ubiquitination and REV1 define temporally distinct mechanisms for controlling translesion synthesis in the avian cell line DT40. Mol. Cell *30*, 519–529.

Eliene Albers, A., Sbroggiò, M., Pladevall-Morera, D., Martin-Gonzalez, J., Hickson, I.D., and Lopez-Contreras Correspondence, A.J. (2018). Loss of PICH Results in Chromosomal Instability, p53 Activation, and Embryonic Lethality. CellReports *24*, 3274–3284.

Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013). Suppression of Gene Expression by G-Quadruplexes in Open Reading Frames Depends on G-Quadruplex Stability. Angew. Chemie *125*, 5632–5636.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R.., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Curr. Biol. *11*, 1183–1187.

Falk, M., Lukášová, E., and Kozubek, S. (2008). Chromatin structure influences the sensitivity of DNA to γ-radiation. Biochim. Biophys. Acta - Mol. Cell Res. *1783*, 2398–2414.
Fang, G., and Cech, T.R. (1993). The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. Cell *74*, 875–885.

Fay, M.M., Lyons, S.M., and Ivanov, P. (2017). RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. J. Mol. Biol. *429*, 2127–2147.

Fernando, P., Kelly, J.F., Balazsi, K., Slack, R.S., and Megeney, L.A. (2002). Caspase 3 activity is required for skeletal muscle differentiation. Proc. Natl. Acad. Sci. *99*, 11025–

11030.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398–402.

Fujita, J., Crane, A.M., Souza, M.K., Dejosez, M., Kyba, M., Flavell, R.A., Thomson, J.A., and Zwaka, T.P. (2008). Caspase activity mediates the differentiation of embryonic stem cells. Cell Stem Cell *2*, 595–601.

Gardner, R.L., and Brook, F.A. (1997). Reflections on the biology of embryonic stem (ES) cells. Int. J. Dev. Biol. *41*, 235–243.

Ge, X.Q., Jackson, D.A., and Blow, J.J. (2007). Dormant origins licensed by excess Mcm2 7 are required for human cells to survive replicative stress. Genes Dev. *21*, 3331–3341.

Ge, X.Q., Han, J., Cheng, E.-C., Yamaguchi, S., Shima, N., Thomas, J.-L., and Lin, H. (2015). Embryonic Stem Cells License a High Level of Dormant Origins to Protect the Genome against Replication Stress. Stem Cell Reports *5*, 185–194.

Gellert, M., Lipsett, M.N., and Davies, D.R. (1962). Helix formation by guanylic acid. Proc. Natl. Acad. Sci. U. S. A. *48*, 2013–2018.

Ghiasi, P., Hosseinkhani, S., Ansari, H., Aghdami, N., Balalaei, S., Pahlavan, S., and Baharvand, H. (2018). Reversible permeabilization of the mitochondrial membrane promotes human cardiomyocyte differentiation from embryonic stem cells. J. Cell. Physiol.

Giraldo, R., and Rhodes, D. (1994). The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. EMBO J. *13*, 2411–2420.

Gonzales, K.A.U.A.U., Liang, H., Lim, Y.-S., Chan, Y.-S., Yeo, J.-C., Tan, C.-P., Gao, B., Le, B., Tan, Z.-Y., Low, K.-Y., et al. (2015). Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. Cell *162*, 564–579.

Graffmann, N., Ncube, A., Wruck, W., and Adjaye, J. (2018). Cell fate decisions of human iPSC-derived bipotential hepatoblasts depend on cell density. PLoS One *13*, e0200416.

Greber, B., Wu, G., Bernemann, C., Joo, J.Y., Han, D.W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P., et al. (2010). Conserved and Divergent Roles of FGF Signaling in Mouse Epiblast Stem Cells and Human Embryonic Stem Cells. Cell Stem Cell *6*, 215–226.

Gu, B., and Zhu, W.-G. (2012). Surf the post-translational modification network of p53 regulation. Int. J. Biol. Sci. *8*, 672–684.

Guilbaud, G., Murat, P., Recolin, B., Campbell, B.C., Maiter, A., Sale, J.E., and Balasubramanian, S. (2017). Local epigenetic reprogramming induced by G-quadruplex ligands. Nat. Chem. *9*, 1110–1117.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development

136, 1063–1069.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., et al. (2014). The DNA methylation landscape of human early embryos. Nature *511*, 606–610.
Gurdon, J.B. (1962). The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. Development *10*.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

Hayashi, K., Lopes, S.M.C. d. S., Tang, F., and Surani, M.A. (2008). Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. Cell Stem Cell *3*, 391–401.

Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004). Loop-Length-Dependent Folding of G-Quadruplexes. J. Am. Chem. Soc. *126*, 16405–16415.

Hebra F, Fagge CH, K.M. (1874). No Title. Dis. Ski. Incl. Exanthemata *3*.
Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.-S., and Johnson, F.B. (2008). Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic Acids Res. *36*, 144–156.

Hirano, T., and Tamae, K. (2012). Differentiation of Embryonic Stem Cells and Oxidative DNA Damage / DNA Repair Systems. J. Stem Cell Res. Ther. *01*, 1–5.

Hirst, C.E., Ng, E.S., Azzola, L., Voss, A.K., Thomas, T., Stanley, E.G., and Elefanty, A.G. (2006). Transcriptional profiling of mouse and human ES cells identifies SLAIN1, a novel stem cell gene. Dev. Biol. *293*, 90–103.
Hoeijmakers, J.H.J. (2001). Genome maintenance mechanisms for preventing cancer. Nature *411*, 366–374.

Hon, J., Martínek, T., Zendulka, J., and Lexa, M. (2017). pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. Bioinformatics *33*, 3373–3379.

Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature *326*, 292–295.

Huang, W.-C., Tseng, T.-Y., Chen, Y.-T., Chang, C.-C., Wang, Z.-F., Wang, C.-L., Hsu, T.-N., Li, P.-T., Chen, C.-T., Lin, J.-J., et al. (2015). Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. Nucleic Acids Res. *43*, 10102–10113.

Huber, M.D., Lee, D.C., and Maizels, N. (2002). G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. Nucleic Acids Res. *30*, 3954–3961.

Huber, M.D., Duquette, M.L., Shiels, J.C., and Maizels, N. (2006). A conserved G4 DNA binding domain in RecQ family helicases. J. Mol. Biol. *358*, 1071–1080.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916.

ter Huurne, M., Chappell, J., Dalton, S., and Stunnenberg, H.G. (2017). Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. Cell Stem Cell *21*, 449-455.e4.

Hwang, Y., Futran, M., Hidalgo, D., Iyer, D.R., Rhind, N., and Socolovsky, M. (2016). Global Increase in Replication Fork Speed during a p57KIP2-Regulated Erythroid Cell Fate Switch. Blood *128*, 698.

James, D., Noggle, S.A., Swigut, T., and Brivanlou, A.H. (2006). Contribution of human embryonic stem cells to mouse blastocysts. Dev. Biol. *295*, 90–102.

Jansen, J.G., Tsaalbi-Shtylik, A., Hendriks, G., Gali, H., Hendel, A., Johansson, F., Erixon, K., Livneh, Z., Mullenders, L.H.F., Haracska, L., et al. (2009). Separate Domains of Rev1 Mediate Two Modes of DNA Damage Bypass in Mammalian Cells. Mol. Cell. Biol. *29*, 3113–3123.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant–containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. Nat. Genet. *41*, 941–945.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-.). *337*, 816–821.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nat. Rev. Mol. Cell Biol. *7*, 335–346.

Jones, P.A., and Takai, D. (2001). The Role of DNA Methylation in Mammalian Epigenetics. Science (80-.). *293*, 1068–1070.

Jones, S.N., Roe, A.E., Donehower, L.A., and Bradley, A. (1995). Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. Nature *378*, 206–208.

Kang, J., Soog Lee, M., and Gorenstein, D.G. (2005). The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: Application to in vitro combinatorial selection of aptamers. J. Biochem. Biophys. Methods *64*, 147–151.

Kareta, M.S., Sage, J., and Wernig, M. (2015). Crosstalk between stem cell and cell cycle machineries. Curr. Opin. Cell Biol. *37*, 68–74.

Kazemier, H.G., Paeschke, K., and Lansdorp, P.M. (2017). Guanine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. Nucleic Acids Res. *45*, 5913–5919.

Kempf, H., Olmer, R., Haase, A., Franke, A., Bolesani, E., Schwanke, K., Robles-Diaz, D., Coffee, M., Göhring, G., Dräger, G., et al. (2016). Bulk cell density and Wnt/TGFbeta signalling regulate mesendodermal patterning of human pluripotent stem cells. Nat. Commun.

7, 13602.

Kennedy, R.D., and D'Andrea, A.D. (2005). The Fanconi Anemia/BRCA pathway: new faces in the crowd. Genes Dev. *19*, 2925–2940.

Khanna., K.K., Keating, K.E., Kozlov, S., Scott, S., Gatei, M., Hobson, K., Taya, Y., Gabrielli, B., Chan, D., Lees-Miller, S.P., et al. (1998). ATM associates with and phosphorylates p53: mapping the region of interaction. Nat. Genet. *20*, 398–400.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Kimura, T., Kawai, K., Fujitsuka, M., and Majima, T. (2007). Monitoring G-quadruplex structures and G-quadruplex–ligand complex using 2-aminopurine modified oligonucleotides. Tetrahedron *63*, 3585–3590.

Kishi, Y., Fujii, Y., Hirabayashi, Y., and Gotoh, Y. (2012). HMGA regulates the global chromatin state and neurogenic potential in neocortical precursor cells. Nat. Neurosci. *15*, 1127–1133.

Kiyonari, H., Kaneko, M., Abe, S., and Aizawa, S. (2010). Three inhibitors of FGF receptor, ERK, and GSK3 establishes germline-competent embryonic stem cells of C57BL/6N mouse strain with high efficiency and stability. Genesis *48*, NA-NA.

de Klein, A., Muijtjens, M., van Os, R., Verhoeven, Y., Smit, B., Carr, A.M., Lehmann, A.R., and Hoeijmakers, J.H. (2000). Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice. Curr. Biol. *10*, 479–482.

Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C.H., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Bühler, M., Liu, P., et al. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell *17*, 471–485.

De Koning, L., Corpet, A., Haber, J.E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. Nat. Struct. Mol. Biol. *14*, 997–1007.

Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D.B., Cuppen, E., and Tijsterman, M. (2008). Mutagenic Capacity of Endogenous G4 DNA Underlies Genome Instability in FANCJ-Defective C. elegans. Curr. Biol. *18*, 900–905.

Kruse, J.-P., and Gu, W. (2009). Modes of p53 Regulation. Cell *137*, 609–622.
Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat. Chem. Biol. *3*, 218–221.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895–2902.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of

histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lam, A.Q., Freedman, B.S., Morizane, R., Lerou, P.H., Valerius, M.T., and Bonventre, J. V (2014). Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. J. Am. Soc. Nephrol. *25*, 1211–1225.

Lancaster, M.A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501*, 373–379.

Lane, D.P. (1992). p53, guardian of the genome. Nature *358*, 15–16.
Le, D.D., Di Antonio, M., Chan, L.K.M., and Balasubramanian, S. (2015). G-quadruplex ligands exhibit differential G-tetrad selectivity. Chem. Commun. *51*, 8048–8050.

Lee, J.-H., and Paull, T.T. (2005). ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex. Science (80-.). *308*, 551–554.

Lee, K.-H., Li, M., Michalowski, A.M., Zhang, X., Liao, H., Chen, L., Xu, Y., Wu, X., and Huang, J. (2010). A genomewide study identifies the Wnt signaling pathway as a major target of p53 in murine embryonic stem cells. Proc. Natl. Acad. Sci. *107*, 69–74.

Lehmann, A.R. (1972). Postreplication repair of DNA in ultraviolet-irradiated mammalian cells. J. Mol. Biol. *66*, 319–337.

Lerner, L.K., Sale, J.E., Lerner, L.K., and Sale, J.E. (2019). Replication of G Quadruplex DNA. Genes 2019, Vol. 10, Page 95 *10*, 95.

Li, A., Yu, Y., Lee, S.-C., Ishibashi, T., Lees-Miller, S.P., and Ausió, J. (2010). Phosphorylation of Histone H2A.X by DNA-dependent Protein Kinase Is Not Affected by Core Histone Acetylation, but It Alters Nucleosome Stability and Histone H1 Binding. J. Biol. Chem. *285*, 17778–17788.

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.-L., et al. (2008). Germline Competent Embryonic Stem Cells Derived from Rat Blastocysts. Cell *135*, 1299–1310.

Li, Y., Geyer, C.R., and Sen, D. (1996). Recognition of anionic porphyrins by DNA aptamers. Biochemistry *35*, 6911–6922.

Lieber, M.R. (2008). The Mechanism of Human Nonhomologous DNA End Joining. J. Biol. Chem. *283*, 1–5.

Lim, D.S., and Hasty, P. (1996). A mutation in mouse rad51 results in an early embryonic lethal that is suppressed by a mutation in p53. Mol. Cell. Biol. *16*, 7133–7143.

Lin, T., Chao, C., Saito, S., Mazur, S.J., Murphy, M.E., Appella, E., and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nat. Cell Biol. *7*, 165–171.

Lindahl, T., and Barnes, D.E. (2000). Repair of endogenous DNA damage. Cold Spring Harb. Symp. Quant. Biol. *65*, 127–133.

Lindahl, T., and Wood, R.D. (1999). Quality control by DNA repair. Science *286*, 1897–1905.

Liu, Q., Guntuku, S., Cui, X.S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A., et al. (2000). Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Genes Dev. *14*, 1448–1459.

London, T.B.C., Barber, L.J., Mosedale, G., Kelly, G.P., Balasubramanian, S., Hickson, I.D., Boulton, S.J., and Hiom, K. (2008). FANCJ Is a Structure-specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts. J. Biol. Chem. *283*, 36132–36139.

Lopes, M., Foiani, M., and Sogo, J.M. (2006). Multiple Mechanisms Control Chromosome Integrity after Replication Fork Uncoupling and Restart at Irreparable UV Lesions. Mol. Cell *21*, 15–27.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lutzker, S.G., and Levine, A.J. (1996). A functionally inactive p53 protein in teratocarcinoma cells is activated by either DNA damage or cellular differentiation. Nat. Med. *2*, 804–810.

Lyons, S.M., Gudanis, D., Coyne, S.M., Gdaniec, Z., and Ivanov, P. (2017). Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. Nat. Commun. *8*, 1127.

Madabhushi, R., Gao, F., Pfenning, A.R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T.X., Yamakawa, H., Pao, P.-C., et al. (2015). Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. Cell *161*, 1592–1605.

Maizels, N., and Gray, L.T. (2013). The G4 Genome. PLoS Genet. *9*, e1003468.
Mak, T.W., Hauck, L., Grothe, D., and Billia, F. (2017). p53 regulates the cardiac transcriptome. Proc. Natl. Acad. Sci. U. S. A. *114*, 2331–2336.

Malashicheva, A., Afanassieff, M., Godet, M., Tapponnier, Y., Petit, M., Coronado, D., Savatier, P., Bourillot, P.-Y., Bernat, A., Iacone, R., et al. (2012). A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. Stem Cell Res. *10*, 118–131.

Marcel, V., Tran, P.L.T., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.-P., Hall, J., Mergny, J.-L., Hainaut, P., and Van Dyck, E. (2011). G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. Carcinogenesis *32*, 271–278.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nat. Rev. Genet. *11*, 285–296.

Marks, H., and Stunnenberg, H.G. (2014). Transcription regulation and chromatin structure in the pluripotent ground state. Biochim. Biophys. Acta - Gene Regul. Mech. *1839*, 129–137.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J.,

Kranz, A., Francis Stewart, A., Smith, A., et al. (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell *149*, 590–604.

Martín-Caballero, J., Flores, J.M., García-Palencia, P., and Serrano, M. (2001). Tumor susceptibility of p21(Waf1/Cip1)-deficient mice. Cancer Res. *61*, 6234–6238.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78*, 7634–7638.

Matson, J.P., Dumitru, R., Coryell, P., Baxley, R.M., Chen, W., Twaroski, K., Webber, B.R., Tolar, J., Bielinsky, A.-K., Purvis, J.E., et al. (2017). Rapid DNA replication origin licensing protects stem cell pluripotency. Elife *6*, e30473.

Matsuoka, S., Huang, M., and Elledge, S.J. (1998). Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. Science *282*, 1893–1897.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., et al. (2007). ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. Science (80-.). *316*, 1160–1166.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. *24*, 529–538.

Meier, A., Fiegler, H., Muñoz, P., Ellis, P., Rigler, D., Langford, C., Blasco, M.A., Carter, N., and Jackson, S.P. (2007). Spreading of mammalian DNA-damage response factors studied by ChIP-chip at damaged telomeres. EMBO J. *26*, 2707–2718.

Menendez, L., Yatskievych, T.A., Antin, P.B., and Dalton, S. (2011). Wnt signaling and a Smad pathway blockade direct the differentiation of human pluripotent stem cells to multipotent neural crest cells. Proc. Natl. Acad. Sci. *108*, 19240–19245.

Mergny, J.-L., Riou, J.-F., Mailliet, P., Teulade-Fichou, M.-P., and Gilson, E. (2002). Natural and pharmacological regulation of telomerase. Nucleic Acids Res. *30*, 839–865.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

Messmer, T., von Meyenn, F., Savino, A., Santos, F., Mohammed, H., Lun, A.T.L., Marioni, J.C., and Reik, W. (2019). Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. Cell Rep. *26*, 815-824.e4.

Mohaghegh, P., Karow, J.K., Brosh, R.M., Bohr, V.A., and Hickson, I.D. (2001). The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. Nucleic Acids Res. *29*, 2843–2849.

Momcilovic, O., Knobloch, L., Fornsaglio, J., Varum, S., Easley, C., and Schatten, G. (2010). DNA Damage Responses in Human Induced Pluripotent Stem Cells and Embryonic Stem Cells. PLoS One *5*, e13410.

Moruno-Manchon, J.F., Koellhoffer, E.C., Gopakumar, J., Hambarde, S., Kim, N.,

McCullough, L.D., and Tsvetkov, A.S. (2017). The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and downregulates transcription of Brca1 in neurons. Aging (Albany NY) *9*, 1957–1970.

Mukundan, V.T., and Phan, A.T. (2013). Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. J. Am. Chem. Soc. *135*, 5017–5028.

Murat, P., and Balasubramanian, S. (2014). Existence and consequences of G-quadruplex structures in DNA. Curr. Opin. Genet. Dev. *25*, 22–29.

Murphy, C.L., and Polak, J.M. (2002). Differentiating embryonic stem cells: GAPDH, but neither HPRT nor beta-tubulin is suitable as an internal standard for measuring RNA levels. Tissue Eng. *8*, 551–559.

Nakken, S., Rognes, T., and Hovig, E. (2009). The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. Nucleic Acids Res. *37*, 5749–5756.

Nazarov, I.B., Smirnova, A.N., Krutilina, R.I., Svetlova, M.P., Solovjeva, L. V, Nikiforov, A.A., Oei, S.-L., Zalenskaya, I.A., Yau, P.M., Bradbury, E.M., et al. (2003). Dephosphorylation of histone gamma-H2AX during repair of DNA double-strand breaks in mammalian cells and its inhibition by calyculin A. Radiat. Res. *160*, 309–317.

Neganova, I., Vilella, F., Atkinson, S.P., Lloret, M., Passos, J.F., von Zglinicki, T., O'Connor, J.-E., Burks, D., Jones, R., Armstrong, L., et al. (2011). An Important Role for CDK2 in G1 to S Checkpoint Activation and DNA Damage Response in Human Embryonic Stem Cells. Stem Cells *29*, 651–659.

Neidle, S., and Parkinson, G. (2002). Telomere maintenance as a target for anticancer drug discovery. Nat. Rev. Drug Discov. *1*, 383–393.

Nguyen, Q.H., Lukowski, S.W., Chiu, H.S., Senabouth, A., Bruxner, T.J.C., Christ, A.N., Palpant, N.J., and Powell, J.E. (2018). Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. Genome Res. *28*, 1053–1066.

Nichols, J., and Smith, A. (2009). Naive and Primed Pluripotent States. Cell Stem Cell *4*, 487–492.

Nichols, J., Jones, K., Phillips, J.M., Newland, S.A., Roode, M., Mansfield, W., Smith, A., and Cooke, A. (2009). Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. Nat. Med. *15*, 814–818.

Nicoludis, J.M., Barrett, S.P., Mergny, J.-L., and Yatsunyk, L.A. (2012a). Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. Nucleic Acids Res. *40*, 5432–5447.

Nicoludis, J.M., Barrett, S.P., Mergny, J.-L., and Yatsunyk, L.A. (2012b). Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. Nucleic Acids Res. *40*, 5432–5447.

Nielsen, C.F., and Hickson, I.D. (2016). PICH promotes mitotic chromosome segregation: Identification of a novel role in rDNA disjunction. Cell Cycle *15*, 2704–2711.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.

Norris, D.P., Patel, D., Kay, G.F., Penny, G.D., Brockdorff, N., Sheardown, S.A., and Rastan, S. (1994). Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell *77*, 41–51.

de Oca Luna, R.M., Wagner, D.S., and Lozano, G. (1995). Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53. Nature *378*, 203–206.

Oganesian, L., and Bryan, T.M. (2007). Physiological relevance of telomeric G-quadruplex formation: a potential drug target. BioEssays *29*, 155–165.

Oka, S., Hayashi, M., Taguchi, K., Hidaka, M., Tsuzuki, T., and Sekiguchi, M. (2019). ROS control in human iPS cells reveals early events in spontaneous carcinogenesis. Carcinogenesis.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Oliner, J.D., Kinzler, K.W., Meltzer, P.S., George, D.L., and Vogelstein, B. (1992). Amplification of a gene encoding a p53-associated protein in human sarcomas. Nature *358*, 80–83.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315.

Pacek, M., and Walter, J.C. (2004). A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. EMBO J. *23*, 3667–3676.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H.J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. Nat. Struct. Mol. Biol. *12*, 847–854.

Pal, R., Mamidi, M.K., Das, A.K., and Bhonde, R. (2012). Diverse effects of dimethyl sulfoxide (DMSO) on the differentiation potential of human embryonic stem cells. Arch. Toxicol. *86*, 651–661.

Papadopoulou, C., Guilbaud, G., Schiavone, D., and Sale, J.E. (2015). Nucleotide Pool Depletion Induces G-Quadruplex-Dependent Perturbation of Gene Expression. Cell Rep. *13*, 2491–2503.

Paria, B.C., Lim, H., Wang, X.-N., Liehr, J., Das, S.K., and Dey, S.K. (1998). Coordination of Differential Effects of Primary Estrogen and Catecholestrogen on Two Distinct Targets Mediates Embryo Implantation in the Mouse [1]. Endocrinology *139*, 5235–5246.

Paria, B.C., Reese, J., Das, S.K., and Dey, S.K. (2002). Deciphering the Cross-Talk of Implantation: Advances and Challenges. Science (80-.). *296*, 2185–2188.

Park, I.-H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Patel, P., and Hosur, R. V. (1999). NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by Saccharomyces cerevisiae telomere repeats. Nucleic Acids Res. *27*, 2457–2464.

Patel, D.J., Phan, A.T., and Kuryavyi, V. (2007). Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic Acids Res. *35*, 7429–7455.

Pauklin, S., and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. Cell *155*, 135–147.

Pauklin, S., Pedersen, R.A., and Vallier, L. (2011). Mouse pluripotent stem cells at a glance. J. Cell Sci. *124*, 3727–3732.

Pauklin, S., Madrigal, P., Bertero, A., and Vallier, L. (2016). Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. Genes Dev. *30*, 421–433.

Payer, B., Chuva de Sousa Lopes, S.M., Barton, S.C., Lee, C., Saitou, M., and Surani, M.A. (2006). Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. Genes. (New York, N.Y. 2000) *44*, 75–83.

Piccinin, S., Tonin, E., Sessa, S., Demontis, S., Rossi, S., Pecciarini, L., Zanatta, L., Pivetta, F., Grizzo, A., Sonego, M., et al. (2012). A "Twist box" Code of p53 Inactivation: Twist box:p53 Interaction Promotes p53 Degradation. Cancer Cell *22*, 404–415.

Pietras, Z., Wojcik, M.A., Borowski, L.S., Szewczyk, M., Kulinski, T.M., Cysewski, D., Stepien, P.P., Dziembowski, A., and Szczesny, R.J. (2018). Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. Nat. Commun. *9*, 2558.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat. Biotechnol. *28*, 848–855.

Rastan, S., and Robertson, E.J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. Development *90*.

Redman, J.E. (2007). Surface plasmon resonance for probing quadruplex folding and interactions with proteins and small molecules. Methods *43*, 302–312.

Reinhardt, H.C., Aslanian, A.S., Lees, J.A., and Yaffe, M.B. (2007). p53-Deficient Cells Rely on ATM- and ATR-Mediated Checkpoint Signaling through the p38MAPK/MK2 Pathway for Survival after DNA Damage. Cancer Cell *11*, 175–189.

Reverón-Gómez, N., González-Aguilera, C., Stewart-Morgan, K.R., Petryk, N., Flury, V., Graziano, S., Johansen, J.V., Jakobsen, J.S., Alabert, C., and Groth, A. (2018). Accurate

Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. Mol. Cell *72*, 239-249.e5.

Rey, L., Sidorova, J.M., Puget, N., Boudsocq, F., Biard, D.S.F., Monnat, R.J., Cazaux, C., and Hoffmann, J.-S. (2009). Human DNA Polymerase η Is Required for Common Fragile Site Stability during Unperturbed DNA Replication. Mol. Cell. Biol. *29*, 3344–3354.

Rhodes, D., and Lipps, H.J. (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. *43*, 8627–8637.

Ribeyre, C., Lopes, J., Boulé, J.-B., Piazza, A., Guédin, A., Zakian, V.A., Mergny, J.-L., and Nicolas, A. (2009). The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. PLoS Genet. *5*, e1000475.
Rigo, R., Bianco, S., Musetti, C., Palumbo, M., and Sissi, C. (2016). Molecular Basis for Differential Recognition of G-Quadruplex versus Double-Helix DNA by Bis-Phenanthroline Metal Complexes. ChemMedChem *11*, 1762–1769.

Robert, M.-F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I.C., Barsalou, A., and MacLeod, A.R. (2003). DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. Nat. Genet. *33*, 61–65.

Rodriguez, R., Müller, S., Yeoman, J.A., Trentesaux, C., Riou, J.-F., and Balasubramanian, S. (2008). A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. J. Am. Chem. Soc. *130*, 15758–15759.

Rodriguez, R., Miller, K.M., Forment, J. V, Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. *8*, 301–310.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat. Struct. Mol. Biol. *19*, 1155–1160.

Rouhani, F., Kumasaka, N., de Brito, M.C., Bradley, A., Vallier, L., and Gaffney, D. (2014). Genetic Background Drives Transcriptional Variation in Human Induced Pluripotent Stem Cells. PLoS Genet. *10*, e1004432.

Safa, L., Gueddouda, N.M., Thiébaut, F., Delagoutte, E., Petruseva, I., Lavrik, O., Mendoza, O., Bourdoncle, A., Alberti, P., Riou, J.-F., et al. (2016). 5′ to 3′ Unfolding Directionality of DNA Secondary Structures by Replication Protein A. J. Biol. Chem. *291*, 21246–21256.

Sahakyan, A.B., Chambers, V.S., Marsico, G., Santner, T., Di Antonio, M., and Balasubramanian, S. (2017). Machine learning model for sequence-driven DNA G-quadruplex formation. Sci. Rep. *7*, 14535.

Sakano, K., Oikawa, S., Hasegawa, K., and Kawanishi, S. (2001). Hydroxyurea induces site-specific DNA damage via formation of hydrogen peroxide and nitric oxide. Jpn. J. Cancer Res. *92*, 1166–1174.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of Eukaryotic Homologous

Recombination. Annu. Rev. Biochem. *77*, 229–257.

Sanders, C.M. (2010). Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. Biochem. J. *430*, 119–128.

Sarkies, P., Reams, C., Simpson, L.J., and Sale, J.E. (2010). Epigenetic instability due to defective replication of structured DNA. Mol. Cell *40*, 703–713.

Sarkies, P., Murat, P., Phillips, L.G., Patel, K.J., Balasubramanian, S., and Sale, J.E. (2012). FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. Nucleic Acids Res. *40*, 1485–1498.

Savatier, P., Lapillonne, H., van Grunsven, L.A., Rudkin, B.B., and Samarut, J. (1996). Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. Oncogene *12*, 309–322.

Savic, V., Yin, B., Maas, N.L., Bredemeyer, A.L., Carpenter, A.C., Helmink, B.A., Yang-Iott, K.S., Sleckman, B.P., and Bassing, C.H. (2009). Formation of Dynamic γ-H2AX Domains along Broken DNA Strands Is Distinctly Regulated by ATM and MDC1 and Dependent upon H2AX Densities in Chromatin. Mol. Cell *34*, 298–310.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J., and Pluckthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc. Natl. Acad. Sci. *98*, 8572–8577.

Schiavone, D., Guilbaud, G., Murat, P., Papadopoulou, C., Sarkies, P., Prioleau, M.-N., Balasubramanian, S., and Sale, J.E. (2014). Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. EMBO J. *33*, 2507–2520.

Schiavone, D., Jozwiakowski, S.K.K., Romanello, M., Guilbaud, G., Guilliam, T.A.A., Bailey, L.J.J., Sale, J.E.E., and Doherty, A.J.J. (2016). PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells. Mol. Cell *61*, 161–169.

Schulz, E.G.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The Two Active X Chromosomes in Female ESCs Block Exit from the Pluripotent State by Modulating the ESC Signaling Network. Cell Stem Cell *14*, 203–216.

Sen, D., and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature *334*, 364–366.

Shao, J., Zhou, B., Zhu, L., Qiu, W., Yuan, Y.-C., Xi, B., and Yen, Y. (2004). In vitro characterization of enzymatic properties and inhibition of the p53R2 subunit of human ribonucleotide reductase. Cancer Res. *64*, 1–6.

Sherman, M.H., Bassing, C.H., and Teitell, M.A. (2011). Regulation of cell differentiation by the DNA damage response. Trends Cell Biol. *21*, 312–319.

Shigeta, M., Ohtsuka, S., Nishikawa-Torikai, S., Yamane, M., Fujii, S., Murakami, K., and Niwa, H. (2013). Maintenance of pluripotency in mouse ES cells without Trp53. Sci. Rep. *3*, 2944.

Shklover, J., Weisman-Shomer, P., Yafe, A., and Fry, M. (2010). Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression. Nucleic Acids Res. *38*, 2369–2377.

Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., and Hurley, L.H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc. Natl. Acad. Sci. *99*, 11593–11598.

Siegfried, Z., and Cedar, H. (1997). DNA methylation: a molecular lock. Curr. Biol. *7*, R305-7.

Siino, J.S., Nazarov, I.B., Svetlova, M.P., Solovjeva, L. V, Adamson, R.H., Zalenskaya, I.A., Yau, P.M., Morton Bradbury, E., and Tomilin, N. V (2002). Photobleaching of GFP-labeled H2AX in chromatin: H2AX has low diffusional mobility in the nucleus. Biochem. Biophys. Res. Commun. *297*, 1318–1323.

Silva, J., and Smith, A. (2008). Capturing Pluripotency. Cell *132*, 532–536.

Sim, Y.-J., Kim, M.-S., Nayfeh, A., Yun, Y.-J., Kim, S.-J., Park, K.-T., Kim, C.-H., and Kim, K.-S. (2017). 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. Stem Cell Reports *8*, 1312–1328.

Simonsson, T., Pecinka, P., and Kubista, M. (1998). DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res. *26*, 1167–1172.

Singh, I., Ozturk, N., Cordero, J., Mehta, A., Hasan, D., Cosentino, C., Sebastian, C., Krüger, M., Looso, M., Carraro, G., et al. (2015). High mobility group protein-mediated transcription requires DNA damage marker γ-H2AX. Cell Res. *25*, 837–850.

Smith, A.G. (2001). Embryo-Derived Stem Cells: Of Mice and Men. Annu. Rev. Cell Dev. Biol. *17*, 435–462.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688–690.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611–615.

Soufi, A., and Dalton, S. (2016). Cycling through developmental decisions: how cell cycle dynamics control pluripotency, differentiation and reprogramming. Development *143*, 4301–4311.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Stavridis, M.P., Lunn, J.S., Collins, B.J., and Storey, K.G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. Development *134*, 2889–2894.
Stiff, T., Walker, S.A., Cerosaletti, K., Goodarzi, A.A., Petermann, E., Concannon, P., O'Driscoll, M., and Jeggo, P.A. (2006). ATR-dependent phosphorylation and activation of ATM in response to UV treatment or replication fork stalling. EMBO J. *25*, 5775–5782.

Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J., and Moine, H. (2011). G–quadruplex RNA structure as a signal for neurite mRNA targeting. EMBO Rep. *12*, 697–704.
Sundquist, W.I., and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature *342*, 825–829.

Šviković, S., and Sale, J.E. (2017). The Effects of Replication Stress on S Phase Histone Management and Epigenetic Memory. J. Mol. Biol. *429*, 2011–2029.

Šviković, S., Crisp, A., Tan-Wong, S.M., Guilliam, T.A., Doherty, A.J., Proudfoot, N.J., Guilbaud, G., and Sale, J.E. (2019). R-loop formation during S phase is restricted by PrimPol-mediated repriming. EMBO J.

Szafranski, K., Lehmann, R., Parra, G., Guigo, R., and Glockner, G. (2005). Gene Organization Features in A/T-Rich Organisms. J. Mol. Evol. *60*, 90–98.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872.

Tarsounas, M., and Tijsterman, M. (2013). Genomes and G-Quadruplexes: For Better or for Worse. J. Mol. Biol. *425*, 4782–4789.

Tchieu, J., Zimmer, B., Fattahi, F., Amin, S., Zeltner, N., Chen, S., and Studer, L. (2017). A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. Cell Stem Cell *21*, 399-410.e7.

Teo, A.K.K., Arnold, S.J., Trotter, M.W.B., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. *25*, 238–250.

Teo, A.K.K., Ali, Y., Wong, K.Y., Chipperfield, H., Sadasivam, A., Poobalan, Y., Tan, E.K., Wang, S.T., Abraham, S., Tsuneyoshi, N., et al. (2012). Activin and BMP4 Synergistically Promote Formation of Definitive Endoderm in Human Embryonic Stem Cells. Stem Cells *30*, 631–642.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Thiagarajan, R.D., Morey, R., and Laurent, L.C. (2014). The epigenome in pluripotency and differentiation. Epigenomics *6*, 121–137.

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., and Hearn, J.P. (1995). Isolation of a primate embryonic stem cell line. Proc. Natl. Acad. Sci. U. S. A. *92*, 7844–7848.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Tichy, E.D., Pillai, R., Deng, L., Liang, L., Tischfield, J., Schwemberger, S.J., Babcock, G.F., and Stambrook, P.J. (2010). Mouse Embryonic Stem Cells, but Not Somatic Cells, Predominantly Use Homologous Recombination to Repair Double-Strand DNA Breaks. Stem Cells Dev. *19*, 1699–1711.

Todd, A.K., Johnston, M., and Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. *33*, 2901–2907.

Tosolini, M., Brochard, V., Adenot, P., Chebrout, M., Grillo, G., Navia, V., Beaujean, N., Francastel, C., Bonnet-Garnier, A., and Jouneau, A. (2018). Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. Sci. Rep. *8*, 5776.

Tovy, A., Spiro, A., McCarthy, R., Shipony, Z., Aylon, Y., Allton, K., Ainbinder, E., Furth, N., Tanay, A., Barton, M., et al. (2017). p53 is essential for DNA methylation homeostasis in naïve embryonic stem cells, and its loss promotes clonal heterogeneity. Genes Dev. *31*, 959–972.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909–918.

Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., and Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. Nature *518*, 344–349.

Turinetto, V., Orlando, L., Sanchez-Ripoll, Y., Kumpfmueller, B., Storm, M.P., Porcedda, P., Minieri, V., Saviozzi, S., Accomasso, L., Cibrario Rocchietti, E., et al. (2012). High Basal γH2AX Levels Sustain Self-Renewal of Mouse Embryonic and Induced Pluripotent Stem Cells. Stem Cells *30*, 1414–1423.

Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. EMBO J. *22*, 5612–5621.

Valton, A.-L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J.-F., and Prioleau, M.-N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. *33*, 732–746.

Vitale, I., Manic, G., De Maria, R., Kroemer, G., and Galluzzi, L. (2017). DNA Damage in

Stem Cells. Mol. Cell *66*, 306–319.

Wallingford, J.B., Seufert, D.W., Virta, V.C., and Vize, P.D. (1997). p53 activity is essential for normal development in Xenopus. Curr. Biol. *7*, 747–757.

Wang, H., Wang, H., Powell, S.N., Iliakis, G., and Wang, Y. (2004). ATR Affecting Cell Radiosensitivity Is Dependent on Homologous Recombination Repair but Independent of Nonhomologous End Joining. Cancer Res. *64*, 7139–7143.

Wang, Q., Zou, Y., Nowotschin, S., Kim, S.Y., Li, Q. V, Soh, C.-L., Su, J., Zhang, C., Shu, W., Xi, Q., et al. (2017). The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. Cell Stem Cell *20*, 70–86.

Wang, X., Goodrich, K., Gooding, A., Naeem, H., cell, S.A.-M., and 2017, undefined Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. Elsevier.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. Cell Stem Cell *10*, 440–454.

Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M., and Gustafsson, C.M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. Proc. Natl. Acad. Sci. U. S. A. *107*, 16072–16077.

Ward, I.M., and Chen, J. (2001). Histone H2AX Is Phosphorylated in an ATR-dependent Manner in Response to Replicational Stress. J. Biol. Chem. *276*, 47759–47762.

Watson, J.D., and Crick, F.H.C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature *171*, 737–738.

Weiss, C.N., and Ito, K. (2015). DNA damage: a sensible mediator of the differentiation decision in hematopoietic stem cells and in leukemia. Int. J. Mol. Sci. *16*, 6183–6201.

Welsh, S.J., Dale, A.G., Lombardo, C.M., Valentine, H., de la Fuente, M., Schatzlein, A., and Neidle, S. (2013). Inhibition of the hypoxia-inducible factor pathway by a G-quadruplex binding small molecule. Sci. Rep. *3*, 2799.

Weyemi, U., Redon, C.E., Choudhuri, R., Aziz, T., Maeda, D., Boufraqech, M., Parekh, P.R., Sethi, T.K., Kasoji, M., Abrams, N., et al. (2016a). The histone variant H2A.X is a regulator of the epithelial–mesenchymal transition. Nat. Commun. *7*, 10711.

Weyemi, U., Redon, C.E., Sethi, T.K., Burrell, A.S., Jailwala, P., Kasoji, M., Abrams, N., Merchant, A., and Bonner, W.M. (2016b). Twist1 and Slug mediate H2AX-regulated epithelial-mesenchymal transition in breast cells. Cell Cycle *15*, 2398–2404.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684–687.

Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989). Monovalent cation-induced

structure of telomeric DNA: The G-quartet model. Cell *59*, 871–880.
Winship, P.R. (1989). An lmproved method for directly sequencing PCR-amplified material using dimethyl sulphoxide. Nucleic Acids Res. *17*, 1266–1266.

Woodford, K.J., Howell, R.M., and Usdin, K. (1994). A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. J. Biol. Chem. *269*, 27029–27035.

Wray, J., and Hartmann, C. (2012). WNTing embryonic stem cells. Trends Cell Biol. *22*, 159–168.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat. Cell Biol. *13*, 838–845.

Wu, Y., Shin-ya, K., and Brosh, R.M. (2008). FANCJ Helicase Defective in Fanconia Anemia and Breast Cancer Unwinds G-Quadruplex DNA To Defend Genomic Stability. Mol. Cell. Biol. *28*, 4116–4128.

Xia, J., Chiu, L.-Y., Nehring, R.B., Bravo Núñez, M.A., Mei, Q., Perez, M., Zhai, Y., Fitzgerald, D.M., Pribis, J.P., Wang, Y., et al. (2019). Bacteria-to-Human Protein Networks Reveal Origins of Endogenous DNA Damage. Cell *176*, 127-143.e24.

Yafe, A., Shklover, J., Weisman-Shomer, P., Bengal, E., and Fry, M. (2008). Differential binding of quadruplex structures of muscle-specific genes regulatory sequences by MyoD, MRF4 and myogenin. Nucleic Acids Res. *36*, 3916–3925.

Yagura, T., Makita, K., Yamamoto, H., Menck, C.F.M., and Schuch, A.P. (2011). Biological sensors for solar ultraviolet radiation. Sensors (Basel). *11*, 4277–4294.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281–292.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science (80-.). *318*, 1917–1920.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.-Q., Paschon, D.E., Miranda, E., Ordóñez, A., Hannan, N.R.F., Rouhani, F.J., et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. Nature *478*, 391–394.

Zermati, Y., Garrido, C., Amsellem, S., Fishelson, S., Bouscary, D., Valensi, F., Varet, B., Solary, E., and Hermine, O. (2001). Caspase activation is required for terminal erythroid differentiation. J. Exp. Med. *193*, 247–254.

Zhang, D.-H., Fujimoto, T., Saxena, S., Yu, H.-Q., Miyoshi, D., and Sugimoto, N. (2010).

Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. Biochemistry *49*, 4554–4563.

Zhang, R., Lin, Y., and Zhang, C.-T. (2008). Greglist: a database listing potential G-quadruplex regulated genes. Nucleic Acids Res. *36*, D372–D376.

Zhang, W., Moore, L., and Ji, P. (2011a). Mouse models for cancer research. Chin. J. Cancer *30*, 149.

Zhang, X., Yalcin, S., Lee, D.-F., Yeh, T.-Y.J., Lee, S.-M., Su, J., Mungamuri, S.K., Rimmelé, P., Kennedy, M., Sellers, R., et al. (2011b). FOXO1 is an essential regulator of pluripotency in human embryonic stem cells. Nat. Cell Biol. *13*, 1092–1099.

Zhang, Y., Cooke, M., Panjwani, S., Cao, K., Krauth, B., Ho, P.-Y., Medrzycki, M., Berhe, D.T., Pan, C., McDevitt, T.C., et al. (2012). Histone H1 Depletion Impairs Embryonic Stem Cell Differentiation. PLoS Genet. *8*, e1002691.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., et al. (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol. Cell *63*, 1066–1079.

Zou, L., and Elledge, S.J. (2003). Sensing DNA Damage Through ATRIP Recognition of RPA-ssDNA Complexes. Science (80-.). *300*, 1542–1548.

Zybaĭlov, B.L., Sherpa, M.D., Glazko, G. V, Raney, K.D., and Glazko, V.I. (2013). [G4-quadruplexes and genome instability]. Mol. Biol. (Mosk). *47*, 224–231.